Detecting a signature of adaptive radiation: diversification in Lake Tanganyika catfishes

Claire Rachel Peart

University College London

A thesis submitted for the degree of Doctor of Philosophy

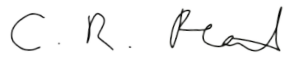
September 2014

UCL

Department of Genetics, Evolution and Environment

Declaration

I, Claire Rachel Peart, confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the thesis.



Claire Rachel Peart
Candidate

Statement of authorship

Chapter 2 has been published in *Molecular Phylogenetics and Evolution* with co-authors Julia Day, Roger Bills and Mark Wilkinson (Appendix 4). I collected and analysed the data and wrote the first draft of the manuscript. Comments from co-authors were incorporated into the text prior to publication.

I took part in a field trip to the Zambian portion of Lake Tanganyika to collect specimens and tissues samples for stable isotope analysis and DNA sequencing used in Chapters 2 - 5. I carried out the majority of the lab work, though some extractions and PCRs for Chapter 2 were carried out with the assistance of a lab technician, Mari-Wyn Burley (UCL). Sanger sequencing was also performed by Mari-Wyn Burley at UCL. Illumina sequencing and RAD-seq library quantification was performed by the FAS Center for Systems Biology, Harvard University. 199 unpublished sequences prepared in the laboratory of Dr Thomas Near are used in Chapter 4. I assembled and aligned all sequence data, and performed all analyses. Stable isotope analyses (Chapter 4) were carried out at the Scottish Universities Environmental Research Centre (SUERC), Life Sciences Mass Spectrometry Facility. I prepared and measured all samples and received training from Jason Newton (SUERC) to run the samples on the mass spectrometer.

Morphological data for 303 *Synodontis* specimens in Chapter 5 were collected jointly by Julia Day and I including specimens from both the Royal Museum of Central Africa, Tervuren and the Natural History Museum, London. I collected data for the remaining 309 specimens in Chapter 5 and all specimens used in Chapter 4 alone.

Abstract

This thesis compares two independent radiations of catfish in Lake Tanganyika, Claroteine and *Synodontis* catfishes, to investigate generalities in patterns and processes of diversification between radiations in an ancient "island-like" environment. The introductory chapter places this work in a theoretical context and explores previous research on taxa from Lake Tanganyika. Chapter two provides the first molecular phylogeny of species in the sub-family Claroteinae from Lake Tanganyika, including additional putative species of the genus *Phyllonemus*. This phylogeny is fossil calibrated to assess when diversification occurred and molecular species delimitation is also performed. Chapter three investigates geographic structure in one species from each radiation, *Lophiobagrus cyclurus* (from the Claroteine radiation) and *Synodontis multipunctatus*. Each species was sampled from multiple localities along the length of Lake Tanganyika and their DNA sequenced using restriction site associated (RAD) sequencing to produce large genomic datasets. These datasets allow the comparison of geographic patterns between taxa and, in the Lophiobagrus dataset (which also includes the closely related species *L. aquilus*), an investigation into the extent of introgression. Chapter four investigates niche partitioning and morphological diversification in both the claroteine and *Synodontis* radiations. These analyses are conducted using a single dated molecular phylogeny containing both of the radiations allowing explicit comparisons. Eco-morphological divergence is assessed using both morphological measurements thought to be ecologically relevant (e.g., size of the eye) and stable isotope ratios of both nitrogen and carbon as a proxy for niche space. While the first three chapters investigate generalities in the patterns and drivers of diversification by comparing two independent radiations in the same "island-like" environment, Chapter fives takes a different approach by placing one of the Lake Tanganyika radiations, *Synodontis*, in its broader phylogenetic context. In this chapter morphological diversification is investigated in the largely riverine continental radiation of *Synodontis* found throughout sub-Saharan Africa.

Table of Contents

Declaration 2

Statement of authorship 2

Abstract 3

Table of Contents 4

Acknowledgements 14

Chapter 1 Lake Tanganyika and adaptive radiation

1.1 Adaptive Radiation - where are we now? 15

1.1.1 The study system 17

1.2 Endemic Radiations of Lake Tanganyika 18

1.2.1 Cichlid fishes 18

1.2.2 Non-cichlid radiations 20

1.2.3 Catfish (siluriforms) in Lake Tanganyika 21

1.2.4 Beyond Lake Tanganyika 23

1.3 Aims for the thesis 24

1.4 References 26

Chapter 2 Nocturnal claroteine catfishes reveal dual colonisation but a single radiation in Lake Tanganyika

2.1 Abstract 34

2.2 Introduction 34

2.2.1 The study system 36

2.3 Methods 37

2.3.1 Sampling 37

2.3.2 DNA extraction, amplification and sequencing 38

2.3.3 Phylogenetic inference 39

2.3.4 Molecular dating 39
2.3.5 Species delimitation 41
2.4 Results 43
2.4.1 Model selection and phylogenetic analysis 43
2.4.2 Estimation of molecular divergence dates 44
2.4.3 Species delimitation 49
2.5 Discussion 50
2.5.1 Origins and colonisation history 50
2.5.2 Diversification of the LT claroteine clade 50
2.5.3 Locally restricted diversity 53
2.5.4 Taxonomy of LT claroteines 54
2.5.5 Conclusions 55
2.6 References 56

Chapter 3 Intralacustrine allopatry: differing geographic structure in two Lake Tanganyika catfish species

3.1 Abstract 61
3.2 Introduction 61
3.3 Methods 65
3.3.1 Sampling 65
3.3.2 Molecular methods 66
3.3.3 Data processing-SNP calling and filtering 67
3.3.4 Population structure 68
3.3.5 Phylogenetic analysis 70
3.3.6 Heterozygosity estimates 70
3.3.7 Introgression tests 70
3.3.8 Relationship with geographic distance 71
3.4 Results 72

3.4.1 RAD-seq characteristics 72
3.4.2 Variable divergence patterns in *L. cyclurus* and *S. multipunctatus* 72
3.4.3 ABBA/BABA SNPs 78
3.4.4 Geographic signature extends to local scales 78
3.5 Discussion 80
3.5.1 Large scale geographic patterns 80
3.5.2 Divergence at local scales 82
3.5.3 Gene flow in *Lophiobagrus* 84
3.5.4 RAD data for studies of recent divergence 85
3.5.5 Conclusions 86
3.6 References 87

Chapter 4 Do sympatric catfish radiations in Lake Tanganyika show patterns of ecological diversification?

4.1 Abstract 97
4.2 Introduction 97
4.3 Methods 101
4.3.1 Taxonomic sampling 101
4.3.2 Phylogenetic tree construction 104
4.3.3 Morphological Analyses 108
4.3.4 Stable isotope analysis 110
4.4 Results 113
4.4.1 Topology and dating of the phylogeny 113
4.4.2 Morphology 116
4.4.3 Trophic partitioning 123
4.5 Discussion 128
4.5.1 Diversification of species in each radiation 128
4.5.2 Resource partitioning 129

4.5.3 Morphological diversity and link with ecology 132
4.5.4 Is there evidence that the radiations are adaptive? 134
4.5.5 Conclusions 136
4.6 References 137

Chapter 5 Morphological divergence at a continental scale in *Synodontis* catfishes

5.1 Abstract 145
5.2 Introduction 146
5.3 Methods 150
5.3.1 Tree preparations 150
5.3.2 Principal components analysis 151
5.3.3 Disparity through time 152
5.3.4 Convergence analysis 153
5.3.5 Geographic Information System (GIS) 154
5.4 Results 156
5.4.1 Principal Component Analysis (PCA) loadings 156
5.4.2 Morphospace 156
5.4.3 Disparity through time 160
5.4.4 SURFACE analysis 162
5.4.5 Geographic distribution of museum collections 167
5.5 Discussion 170
5.5.1 Lake Tanganyika taxa 170
5.5.2 Morphological change through time 170
5.5.3 Evidence for convergent evolution 172
5.5.4 Colonisation of Southern Africa 173
5.5.5 Geographical sampling of Synodontis specimens 173
5.5.6 Conclusions 174

5.6 References 176

Chapter 6

Chapter 6 180

Conclusions 180

6.1 Summary of main results 180

6.2 Lake Tanganyika claroteines 181

6.3 Is there a common pattern of siluriform diversification in Lake Tanganyika? 182

6.4 How does diversification in Lake Tanganyika relate to patterns of diversification at the continental scale? 183

6.5 Future directions 184

6.6 References 187

Appendix 1 – Supplementary materials to Chapter 2 189

Appendix 2 – Supplementary materials to Chapter 3 197

Appendix 3 – Supplementary materials to Chapter 4 203

Appendix 4 - Nocturnal claroteine catfishes reveal dual colonisation but a single radiation in Lake Tanganyika. Molecular Phylogenetics and Evolution, **73**, 119–128 221

List of Figures

Chapter 1

Figure 1 Map of Lake Tanganyika. 17
Figure 2 Dated phylogeny of the LT *Synodontis* 22

Chapter 2

Figure 1 Phylogeny of the Lake Tanganyika claroteines 45
Figure 2 Phylogeny of the Lake Tanganyika claroteines constructed using nuclear data 46
Figure 3 Phylogeny of the Lake Tanganyika claroteines constructed using mitochondrial data. 47
Figure 4 Chronograms of the LT claroteines. 48
Figure 5 Guide tree and results of the BP&P analysis 49

Chapter 3

Figure 1 Map showing sampling locations for *Synodontis multipunctatus* and *Lophiobagrus* species.. 65
Figure 2. RAxML trees and STRUCTURE plots for a) *Lophiobagrus*, b) *S. multipunctatus* 74
Figure 3 PCA plots for (a) *L. cyclurus* and (b) *S. multipunctatus* populations.. 75
Figure 4 Bar charts and boxplots showing heterozygosity in (a) *Lophiobagrus* and (b) *S. multipunctatus* 77
Figure 5 Nominal D Statistics for ABBA-BABA test in *Lophiobagrus*. 79

Chapter 4

Figure 1 Sampling sites in Zambia. 112

Figure 2 Chromogram of the 'Big Africa' clade. 115
Figure 3 Phylogenetically corrected PCA plots of the LT claroteines and *Synodontis* 118
Figure 4 Non-phylogenetically corrected PCA plots for PC axes 1 and 2 for the claroteine and *Synodontis* LT radiations. 119
Figure 5 Disparity through time plots for the claroteine and *Synodontis* LT radiations..... 120
Figure 6 Morphological and stable isotope disparity through time (DTT) plots with chronograms for the *Synodontis* and claroteine LT radiations displayed on the same time scale..... 121
Figure 7 Scatter plots of stable isotope values for the claroteine and *Synodontis* LT radiations showing standard ellipses. 126
Figure 8 Phylogenetic tree of both the claroteine and *Synodontis* showing the carbon and nitrogen values for each species sampled in Zambia.. 127

Chapter 5

Figure 1 Morphospace wheel plot for PC axes 1-3, and PCA plots for PC1 vs axes 2-4 for specimens in the phylogeny..... 158
Figure 2 Morphospace wheel plot for PC axes 1-3, and PCA plots for PC1 vs axes 2-4 for all measured specimens.. 159
Figure 3 Disparity through time plots for a – PC1-4, b – Standard length, c – PC1, d – PC2, e – PC3 and f – PC4 based on the MCC. 161
Figure 4 Results of the SURFACE analysis for PC1-PC4..... 164
Figure 5 Results of the SURFACE analysis for standard length and PC1-PC4..... 165
Figure 6 Histograms showing values of c (the number of shifts towards convergent regimes occupied by multiple lineages) from 99 simulated datasets using the final model from the 'forward' phase of the SURFACE analysis..... 166
Figure 7 a) Sampled range of the taxa suggested to be convergent by SURFACE (clade A and clade B), b) Overlap of predicted distribution of all 11 species in clades A and B 168
Figure 8 a) Predicted climatic range of *Synodontis* (MaxEnt balance threshold) vs actual sampled data b) Specimen data coloured by presence/absence of morphological data for that species in this study at species in this study..... 169

List of Tables

Chapter 2

Table 1 Features of Lake Tanganyika radiations taken from the literature. 51

Chapter 3

Table 1 Fst values for (a) *Lophiobagrus* and (b) *S. multipunctatus* (Index from Reich et al. 2009) 75

Chapter 4

Table 1 MDI for each PC axes using the MCC tree and 1000 trees 122
Table 2 Akaike weights for each PC axes 123
Table 3 Standard area of ellipses using both carbon and nitrogen data. 123
Table 4 Overlap of standard ellipse values. 124

Chapter 5

Table 1 Explanation of bioclimatic indicators 155
Table 2 Results of model fitting for PC axes 1-4 and standard length 160
Table 3 MDI values for the MCC tree and across a subsample of 100 trees 162
Table 4 Calculated parameters from the SURFACE analysis with and without size correction 163

Acknowledgements

Firstly, thanks go to my supervisors, Julia Day (UCL), for enabling me to work on this system and support throughout and Mark Wilkinson (Natural History Museum, London) for always welcoming me at the NHM and valuable advice. Thanks also go to the past and present members of Day lab group at UCL, for their help and friendship along the way: Katherine Brown, Siobhan Cox, Antonia Ford, Jake Morris and Simon Maddock.

This thesis would not have been possible if not for the help of a number collaborators, for which I am very grateful. Kanchon Dasmahapatra (University of York) provided invaluable training and reagents for the generation and analysis of RAD-seq data. During my visits to York thanks go to Neil Rosser and Jake Morris for accepting a temporary flatmate. Thomas Near (Yale University) and his lab group were very welcoming, and provided both guidance and DNA sequences used in this thesis. Jason Newton and the staff at the Scottish Universities Environmental Research Centre (SUERC), Life Sciences Mass Spectrometry Facility were very helpful with training on mass spectroscopy for stable isotope analysis.

For their assistance during the field trip to Zambia thanks go to Roger Bills, Katherine Brown, Austin Brown and staff at the Department of Fisheries, Mpulungu and the Nsumbu Fisheries. In particular, thanks go to Lawrence Makasa and Danny Sinyinza for facilitating the research permits. This thesis was also dependent upon samples collected by Julia Day, Katherine and Austin Brown and George Kazumbe from Kigoma in Tanzania and facilitated by TAFIRI scientists Yohana Budeba, Ben Ngatunga and Amon Shoko. Samples from Burundi were collected by George Kazumbe and thanks go to Saskia Marijnissen for her help in arranging the Burundi collections.

In addition, Roger Bills (South African Institute of Aquatic Biodiversity), John Friel (Cornell University) and Timo Moritz (Meeresmuseum, Katharinenberg) provided important tissue samples. Emmanuel Vreven (RMCA, Tervuren), John Friel (Cornell University) and Barbara Brown (AMNH) facilitated museum collection visit for which I am very appreciative.

Thanks also go to Mari-Wyn Burley for lab support and for assistance in Sanger sequencing and to the FAS Center for Systems Biology, Harvard University for Illumina sequencing support.

For funding, I acknowledge the role of a NERC departmental training grant that funded my studentship (NERC NE/I528169/1), the Systematics Research Fund, a NERC grant-in-kind allowing the use of the Life Sciences Mass Spectrometry Facility at east Kilbride, UCL Graduate school and a Synthesys grant funding the visit to the Royal Museum of Central Africa, Tervuren (with Julia Day). The exchange to the lab group of Thomas Near was as part of the Yale-UCL collaborative exchange programme and funded by a Bogue fellowship for which I am extremely grateful. A National Geographic Society Waitts grant (1341-0) to Julia Day is also acknowledged for funding fieldwork.

Last but certainly not least I'd like to thank all my family and friends for putting up with me.

This thesis is dedicated to my parents for always believing in me and James who has lived through this thesis and read every word- twice.

Chapter 1
Lake Tanganyika and adaptive radiation

1.1 Adaptive Radiation - where are we now?

Adaptive radiation, the diversification of organisms into forms filling different ecological niches, has been used to explain spectacular radiations of organisms and has even been proposed as a mechanism to generate the entire diversity of life (Simpson, 1953). This process has been used to explain both spectacular recent radiations, for example, Darwin's finches (discussed in detail Grant, 1986), Caribbean *Anolis* lizards (discussed in detail Losos, 2009) and the cichlid fishes of East Africa (reviewed in Seehausen, 2006) and also older radiations taking place over greater time scales, for example sauropods (Sander et al., 2004) and multituberculate mammals (Wilson et al., 2012). Adaptive radiation was formally characterised by Schluter (2000) who put forward four criteria to detect adaptive radiation focussing on recent radiations: common ancestry, a correlation between phenotype and environment, traits leading to a fitness advantage in its environment (trait utility), and rapid speciation. Common ancestry is often uncontroversial (monophyly is not required to define an adaptive radiation) and can be elucidated accurately using a multi-gene molecular phylogeny. A phenotype-environment correlation is more difficult to investigate as it frequently depends on field observations of associations between phenotypes and the use of alternate resources or different features of the environment that may not be visually obvious. This correlation can only be tested once the observations have been controlled for non-independence due to phylogenetic relatedness. The criterion of trait utility is necessary to exclude trivial phenotypic associations, and stipulates that traits must actually enhance performance in that environment. This can be tested theoretically, usually by using mechanistic models, for example investigating forces acting on the bills of Darwin's finches (Bowman, 1961) or experimentally, for example, experimental manipulation of flower orientation and spur length in columbines (Fulton and Hodges, 1999).

The theory of adaptive radiation has withstood the test of time with research focussed on identifying adaptive radiations (which may not be as obvious as the famous examples) and investigating the repeatability and predictability of

adaptive radiation. This includes research into situations where radiation has failed to occur despite similar sets of circumstances to those seen in well-studied adaptive radiations, for example in African cichlids (Seehausen, 2006; Wagner et al., 2012a). There is also further work to be done determining the extent of generalisations between adaptive radiations. In large radiations replicated phenotypes are common, for example in the *Anolis* lizards there are not only repeated phenotypes across islands (Losos et al., 1998) but convergence of entire faunas across islands (Mahler et al., 2013). In the East African cichlids there are not only repeated phenotypes between lakes, for example between Lake Tanganyika (LT) and Lake Malawi (Fryer and Iles, 1972; Kocher et al., 1993) but also convergent phenotypes in sympatry, for example in LT (Muschick et al., 2012). In contrast, the pattern in smaller radiations is less clear, for example in the cichlid radiations of the Cameroonian crater lakes Barombi Mbo (11 species) and Bermin (nine species) there are not equivalent species assemblages whereas smaller radiations in Nicaraguan crater lake cichlids do show parallel phenotypes (Elmer et al., 2014).

Recent advances in the field of adaptive radiation have come from expansions in the field of genetics as the revolution in sequencing technology has enabled more genetic data to be collected, for example RAD-seq data now resolves previously unsupported relationships within the phylogeny of Lake Victoria cichlids (Wagner et al., 2013). These sequencing technologies also allow genomic resources to be developed for non-model taxa, expanding the number of radiations that can be studied in depth. This data has allowed for study of the genetic structure of ecological speciation, for example the role of standing genetic variation in the repeated invasions of freshwater conditions by sticklebacks originally from marine habitats (Schluter and Conte, 2009). Investigations into the genetic basis of replicate ecotypes in different radiations have also been conducted revealing broad examples of parallelism and also some local changes (e.g., thick lipped cichlids, Colombo et al., 2013; Manousaki et al., 2013). Other data types have also aided the study of adaptive radiation, for example the growth in geometric morphometrics (reviewed in Adams et al., 2004). Methodological advances in the dating of phylogenetic trees (e.g., Drummond et al., 2006; Drummond and Suchard, 2010), the estimation of species trees (e.g., Heled and Drummond, 2010) and the field of comparative phylogenetics have also aided the

study of adaptive radiation, through, for example, the R packages Geiger (Harmon et al., 2008) and Phytools (Revell, 2012).

1.1.1 The study system

Lake Tanganyika is the oldest, 9-12Ma (Cohen et al., 1993a) and deepest, 1470m (Alin and Cohen, 2003) of the East African Rift Lakes. There is one major outflow from LT, the Lukuga River in the Democratic Republic of Congo which ultimately drains into the Congo River. Multiple rivers and streams drain into LT, the largest volume of water comes from the Ruzizi River at the north of LT, which drains from Lake Kivu. Other rivers of note include the Malagarasi River in Tanzania, which before the formation of LT was connected to the Lukuga River and ultimately the Congo basin, the Kalambo River on the border of Zambia and Tanzania and the Lufubu River at the southern end of the basin (Figure 1). Lake Tanganyika is meromictic with a deeper zone of mixing at the southern end of LT than the northern end with upwelling caused by wind at the southern end (Coulter and Spigel, 1991). The diversity and functioning of LT is at risk from over-fishing (McIntyre et al., 2007; van Zwieten et al., 2002), changing nutrient cycling regimes, particularly from land use change increasing sedimentation (Alin et al., 1999; Bootsma and Hecky, 1993; Cohen et al., 1993b) and from climate change (O'Reilly et al., 2003; Tierney et al., 2010; Verburg and Hecky, 2009).

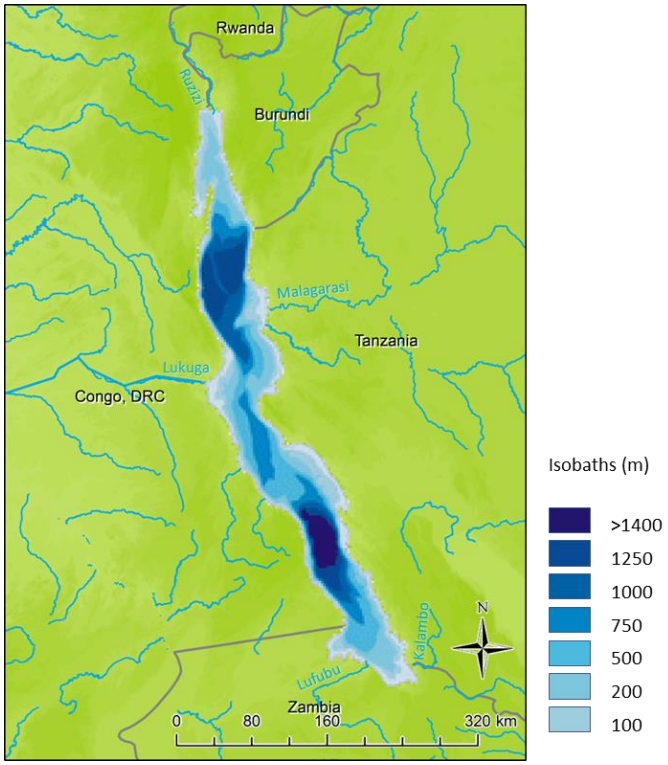


Figure 1 Map of Lake Tanganyika showing major rivers and depth contours, taken from McGlue et al. (2008).

Lake Tanganyika, although undoubtedly ancient, has had a complex geological history with several periods of low lake levels, however, it has not completely desiccated, unlike other large East African rift Lakes such as Lake Malawi or Lake Victoria. Lake Tanganyika began as a collection of shallow lakes (in the central basin) before the northern basin was formed as result of tectonic activity ≈ 7.4 Ma (Cohen et al., 1997). A full lacustrine habitat was achieved in the northern and central basins before the formation of the southern basin 4-2 Ma (Cohen et al., 1997; Klerkx et al., 1998). Strong tectonic activity at 1.1 Ma (Cohen et al., 1997) led to a major lake level decline of around 650-700 m below present levels (Cohen et al., 1997) which led to isolated sub-lakes within the LT basin, by 670 Ka LT was still not fully connected though lake levels had risen (Cohen et al., 1997; Lezzar et al., 1996). Following this there have been repeated periods of lower lake levels related to climate change with lake levels dropping over 400m from their present level (Cohen et al., 1997; McGlue et al., 2008), however bathymetric maps suggest that, at least in the late Pleistocene, fluctuations of this size leave LT as a large and relatively connected lake (McGlue et al., 2008). Lake levels only began to permanently rise around 12Ka following the establishment of a post glacial climate (Scholz et al., 2003).

1.2 Endemic Radiations of Lake Tanganyika

1.2.1 Cichlid fishes

The most intensively studied of the endemic radiations are the LT cichlids, which although less species rich than those found in the other East African Great Lakes show higher phenotypic and ecological diversity (Chakrabarty, 2005). The ages of the largest Tanganyikan cichlid clades, calibrated using fossil data, are within the timeframe of lacustrine conditions in LT, Lampologini 7.29 Ma ± 1.14 (standard deviation), Ectodini 10.8 Ma ± 1.94 and Tropheini 3.02 Ma ± 0.49 (Genner et al., 2007). Older dates were proposed based on dates of Gondwanan fragmentation with a date of 23.53 ± 4.14 Ma proposed for the Ectodini (Genner et al., 2007), however, these dates have been disputed and paleontological evidence plus large molecular phylogenies calibrated with non-cichlid fossils suggest that cichlids originated ≈ 65- 57 Ma, much later than Gondwanan fragmentation (Friedman et al., 2013).

Research on cichlid fishes has focussed on general patterns of diversification (e.g., Day et al., 2008; Salzburger et al., 2005) as well as similarities and differences between different cichlid radiations, for example the same mutation causing blue-shifted rod pigments in cichlid lineages that colonised deep water in both Lake Malawi and Lake Tanganyika (Sugawara et al., 2005). In general LT cichlids are thought to follow the same general pattern of diversification as put forward for Lake Malawi cichlids (Danley and Kocher, 2001) with initial diversification into macrohabitats, followed by diversification driven by ecological specialisation and recent splits influenced by sexual selection (discussed in Salzburger, 2009). Earlier studies have considered the development of the pharyngeal jaw in cichlids as a key morphological innovation facilitating diversification (Liem, 1973). However, the propensity to radiate in lakes seems to be a derived property in African cichlids (Seehausen, 2006), that does not relate to the presence of the pharyngeal jaw or originate at the same time as sexual dimorphism in anal fin or parental care (as suggested by Salzburger et al., 2005). The presence of sexual dichromatism (as a proxy for sexual selection) does correlate with diversification as does depth of the lake, more incident solar radiation and time for diversification (Wagner et al., 2012a).

The relative contributions of different factors to the diversification and co-existence of LT cichlids is still an active area of research. Most of the research on resource partitioning has focussed on dietary segregation (e.g., Sturmbauer et al., 1992), however competition for habitats has been shown to maintain reproductive isolation in two morphs of *Telmatochromis temporalis* (Winkelmann et al., 2014). Competition resulting from resource partitioning is thought to lead to increased specialization and thus morphological and ecological diversification, however in LT cichlids there is considerable convergence in ecomorphology between distantly related species (Muschick et al., 2012) which suggests that species may be able to co-exist by virtue of being very similar to each other. The role of resource partitioning may vary in space and time too as species with specialized morphologies have been shown to feed on a broader range of resources than they are specialized to feed on (Liem, 1980).

Hybridisation, as evidenced by nuclear-mitochondrial discordance, between distantly related species followed from ecological specialization to shell breeding in the lamprologines (Koblmüller et al., 2007). Hybridisation between

two species has also been implicated in the formation of a new species (Salzburger et al., 2002). Lake level fluctuations can also contribute to increased diversity through hybridisation when lake levels rise allowing secondary contact between previously isolated populations, for example in *Tropheus* (Egger et al., 2007).

In LT many rocky shore cichlids are philopatric and do not cross sandy bays showing isolation by distance and microallopatry across small distances, for example there is genetic structure seen across a 7km bay in Zambia in a variety of cichlid species (Sefc et al., 2007). In this case geographical isolation is an important factor driving diversification, however geographic patterns do vary between species, for example in *Tropheus* some mitochondrial lineages are widespread whereas others are restricted (Baric et al., 2003) and some species show little geographic structure including both littoral species (e.g., Meyer et al., 1996) and those found in deeper water (Koblmüller et al., 2014). Geographic isolation can lead to multiple patterns of diversification including cryptic speciation or to the evolution of morphological difference without reproductive isolation.

Sexual selection has also been proposed to play a role in diversification of species, behaviours and phenotypes in LT cichlids. Evolutionary changes in parental care in LT cichlids are dependent on the intensity of sexual selection, using data on sexual dichromatism and dimorphism data as a proxy (Gonzalez-Voyer et al., 2008). There is also evidence for recent speciation between sympatric Tanganyikan colour morphs (Wagner et al., 2012b) suggesting that the factors driving this are on-going. However, in some cases it has been suggested that sexual isolation based on colour morphs evolved as a by-product of allopatric speciation (Zoppoth et al., 2013). The presence of egg dummies on male anal fins is also correlated to the presence of sexual selection and is thought to be an example of evolution via sensory exploitation (Amcoff et al., 2013). In addition to the effect of each separate factor on cichlid diversification it is also important to consider that different factors can combine to influence diversification, for example in Tropheini trophically specialized species were more restricted geographically than a more generalist species (Wagner and McCune, 2009).

1.2.2 Non-cichlid radiations

While the LT cichlid radiations are the best-studied radiations from this system, it is the diversity of endemic radiations that sets LT apart from the other East African

Great Lakes. These include platytelphusid crabs (Marijnissen et al., 2006), multiple gastropod lineages (West and Michel, 2000; Wilson et al., 2004), ostracods (Schön and Martens, 2011), atyid prawns (Fryer, 2006), multiple sponge lineages (Erpenbeck et al., 2011), mastacembelid eels (Brown et al., 2010) and catfishes (discussed below). The presence of multiple radiations allows the patterns and process of radiation in this ancient lake to be investigated, for example it appears that most radiations diversified in full lacustrine conditions, however, it has been suggested that LT acts as a refuge of ancient diversity in thalassoid gastropods (Wilson et al., 2004). The presence of independent radiations allows the effects of external factors (e.g., lake levels change) to be elucidated as they would leave similar signatures in different taxa that occupy the same environments, for example, rocky shore cichlids and mastacembelid eels. There are currently a number of problems in making comparisons between these radiations, for example, different dating methodologies used in each study and the problem of calibrating recent radiations (discussed more fully in Chapter 2). Poor sampling from around LT particularly from the western shore in the Democratic Republic of the Congo and poor taxonomy of these taxa both hinder studies of diversification at the species and population levels. New species have recently been described in the LT mastacembelid eels (Brown et al., 2011), as well as in multiple catfish genera, as discussed below. Within each of these non-cichlid radiations there has been little investigation into the intrinsic factors influencing these radiations and the role of their ecologies and breeding system remains to be studied. Resource partitioning has been investigated in playthelphusid crabs, which show interspecific segregation in depth, substrate type and diet (Marijnissen et al., 2008) though there is some dietary overlap.

1.2.3 Catfishes (siluriforms) in Lake Tanganyika

A variety of catfishes are known from Lake Tanganyika (≈ 34 species from five different families; Bagridae, Clariidae, Claroteidae, Malapteruridae, and Mochokidae, with species from two additional families, Amphilidae and Schilbeidae found in the surrounding waters (Coulter, 1991). Only the subfamily Claroteinae (Claroteidae), and the genera *Synodontis* (Mochokidae) (Day and Wilkinson, 2006; Koblmüller et al., 2006) and *Tanganikallabes* (Clariidae) (Wright and Bailey, 2012) form radiations. *Tanganikallabes* was a formerly monotypic genus that has been expanded to encompass three species *T. mortiauxi*,

T. alboperca and *T. stewarti*, although this radiation has not been dated and the patterns and processes driving diversification in this group remain to be studied. Prior to this thesis, *Synodontis* is the only catfish radiation that has both had its evolutionary relationships established and been dated (Day and Wilkinson, 2006; Day et al., 2009; Day et al., 2013; Koblmüller et al., 2006). There are currently 11 species of *Synodontis* described from Lake Tanganyika (Wright and Page, 2006), though some species designations are controversial (e.g., Van Steenberge et al., 2011; discussion in Chapter 4) and one species, *S. dhonti,* is described from a single specimen. *Synodontis* are diurnal (in contrast to the nocturnal Claroteinae) and are known to exhibit multiple breeding strategies with *S. multipunctatus* a brood parasite (Sato, 1986).

The onset of divergence within the endemic *Synodontis* species in LT has been estimated to be between 7.9 Ma (5.7-10 Ma) (Day et al., 2013), however the age estimate is based on a single fossil calibration and using a different calibration was dated as 5.5 Ma (4.0-7.3 Ma) (Day et al., 2009). The factors influencing this radiation and enabling co-existence of species are largely unknown, indeed it has recently been found that they are Müllerian mimics with selection maintaining their strikingly conserved colour patterns (Wright, 2011) which could potentially be a key innovation that helped to facilitate the radiation. The adaptive character of this radiation has not yet been established.

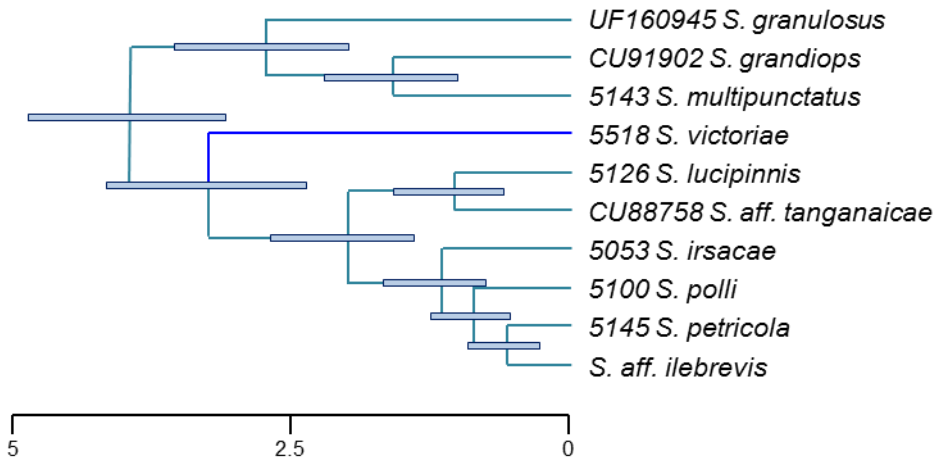


Figure 2 Dated phylogeny of the LT *Synodontis* radiation (edited from Day et al., 2013). Scale bar is in Ma.

In the family Claroteidae there are two sub-families, the Claroteinae and the Auchenoglaninae. The monophyly of this family has been called into question

(Sullivan et al., 2006) though the claroteid taxa that have radiated within LT are all in the sub-family Claroteinae which is monophyletic (Sullivan et al., 2006). In the LT Claroteinae radiation there are four genera, *Chrysichthys* (seven species), *Lophiobagrus* (four species), *Phyllonemus* (three species currently described but is purported to contain nine species, R. Bills *pers. comm.*) and *Bathybagrus* (single species), though the evolutionary relationships within this radiation have not yet been ascertained. There is currently lingering taxonomic confusion based on morphological features with Mo (1991) placing *Chrysichthys grandis, C. graueri, C. platycephalus,* and *C. sianenna* in *Bathybagrus* though this diagnosis has proven controversial and more recent studies have reverted to the classification of Bailey and Stewart (1984) based upon the unique features of *Bathybagrus* (subcutaneous eyes and an absence of nasal and inner mandibular barbels) and concern over the identity of Mo's specimens (e.g., Hardman, 2008). These smaller radiations within the Claroteinae present a fascinating system to compare with *Synodontis* to investigate generalisations in catfish radiations, particularly as some *Lophiobagrus* and *Phyllonemus* species are known to be mouth brooders (Ochi et al., 2000, 2001, 2002).

1.2.4 Beyond Lake Tanganyika

Continental freshwater studies from Africa are very rare, with some areas historically under sampled, including parts of the Congo basin (which is particularly species rich) and Angola. In the studies that do exist, there is evidence that lacustrine radiations of cichlids are more diverse than those seen in riverine systems, however riverine systems are more diverse in *Synodontis* and the Claroteinae.

The LT haplochromine taxa (Tropheni) are sister to some riverine taxa, as well as species flocks from Lake Malawi, Lake Victoria and surrounding lakes suggesting LT may be a source of species diversity in this group, as are some riverine systems, for example, some taxa started to diversify in the upper Congo system then expanded to the north and south (Koblmüller et al., 2008). A similar pattern is also seen in some cichlids in the Lamprologini tribe, in which species are found in both the Congo and Malagarasi River but neither of these riverine clades is ancestral to the clade in LT (Day et al., 2007; Sturmbauer et al., 2010). The importance of lacustrine diversification for cichlids is further highlighted by the hypothesis that the diverse phenotypes of haplochromine cichlids in southern

African rivers originated in the now extinct Pleistocene Lake Makgadikgadi (Joyce et al., 2005). Zambian serranochromine cichlids also show signature of diversifying in Lake Makgadikgadi (Katongo et al., 2007), a hypothesis that has also been proposed for *Synodontis* (Day et al., 2009). In *Synodontis* there is also evidence for diversification in lacustrine environments seeding rivers through the emigration of *S. victoriae*, which resolves within the LT clade but is not found in LT. The full characterisation of the patterns seen in the Claroteinae is hampered by the lack of molecular phylogenies and a lack of sampling, especially in the largest genus *Chrysichthys,* which is most diverse in the Congo basin.

1.3 Aims for the thesis

The overarching aims of this thesis are to compare and contrast the patterns and processes of diversification in two catfish radiations from LT and to place the LT siluriform radiations within the broader evolutionary framework of siluriform diversification in Africa. This thesis provides an integrated study to address these aims and utilises a broad range of datasets including DNA sequences (Sanger sequencing and restriction site associated sequencing), diet proxies, morphology, fossil calibrations and georeferenced museum collections.

In Chapter 2, novel data is used to generate the first multigene molecular phylogeny of the LT Claroteinae including broader taxonomic sampling, in order to resolve the evolutionary relationships in this group (including tests for if the LT radiation is monophyletic), establish the colonisation history of this group and use a fossil calibration to date the claroteine radiation. A multi-gene phylogeny (both nuclear and mitochondrial) will be used in order to investigate if there is introgression as conflicting signals between loci have been found in LT cichlids (Salzburger et al., 2002), but also in other smaller species assemblages in ancient lakes for example the sailfin silversides in the Malili Lakes of Sulawesi (Herder et al., 2006). In addition, the validity of putative new *Phyllonemus* species (R. Bills *pers. comm.*) is assessed.

In Chapter 3, sequencing from throughout the genome (RAD-seq data) is applied for the first time to a non-cichlid lineage within the East African Great Lakes to investigate diversification at an intra-specific level in *Synodontis multipunctatus*,

and the claroteid species *Lophiobagrus cyclurus*. Evidence for phylogeographic structure at both a lake-wide and local scale is investigated in order to assess how geographic distance influences genetic structure in both species.

In Chapter 4, both the LT claroteine and LT *Synodontis* radiations are placed in a common dated phylogenetic context for the first time, and morphological data is used to compare and contrast the timing, mode and the adaptive nature of these two radiations. In addition stable isotope data (carbon and nitrogen isotopes) are used to assess niche partitioning in diets in both the *Synodontis* and claroteine radiations using samples from Zambia.

Finally in Chapter 5, a different approach is taken; instead of comparing multiple radiations within LT, morphological diversification in the *Synodontis* radiation is investigated at its broader phylogenetic context using a published multi-gene molecular phylogeny of the genus, containing ≈ 60% species and incorporating species from throughout Africa (Day et al., 2013). Collection information from museum specimens is also used to assess geographic patterns.

1.4 References

Adams, D. C., Rohlf, F. J., & Slice, D. E. (2004) Geometric morphometrics: Ten years of progress following the "revolution." *Italian Journal of Zoology*, **71**(1), 5–16

Alin, S., & Cohen, A. (2003) Lake-level history of Lake Tanganyika, East Africa, for the past 2500 years based on ostracode-inferred water-depth reconstruction. *Palaeogeography, Palaeoclimatology, Palaeoecology*, **199**(1-2), 31–49

Alin, S. R., Cohen, A. S., Bills, R., Gashagaza, M. M., Michel, E., Tiercelin, J. J., & Ntakimazi, G. (1999) Effects of Landscape Disturbance on Animal Communities in Lake Tanganyika, East Africa. *Conservation Biology*, **13**(5), 1017–1033.

Amcoff, M., Gonzalez-Voyer, A., & Kolm, N. (2013) Evolution of egg dummies in Tanganyikan cichlid fishes: the roles of parental care and sexual selection. *Journal of Evolutionary Biology*, **26**(11), 2369–2382

Bailey, R. M., & Stewart, D. J. (1984) Bagrid catfishes from Lake Tanganyika, with a key and descriptions of new taxa. *Miscellaneous Publications Museum of Zoology, University of Michigan*, **168**

Baric, S., Salzburger, W., & Sturmbauer, C. (2003) Phylogeography and evolution of the Tanganyikan cichlid genus Tropheus based upon mitochondrial DNA sequences. *Journal of Molecular Evolution*, **56**(1), 54–68

Bills, R. I., & Lundberg, J. G. (n.d.) Descriptions of new species of Phyllonemus catfish (Siluriform: Claroteidae) from Lake Tanganyika, Africa. *In Prep*

Bootsma, H. A., & Hecky, R. E. (1993) Conservation of the African Great lakes: A Limnological Perspective. *Conservation Biology*, **7**(3), 644–656

Bowman, R. I. (1961) Morphological differentiation and adaptation in the Galapagos finches. *University of California Publications in Zoology*, **58**, 1–302

Brown, K. J., Britz, R., Bills, R., Rüber, L., & Day, J. J. (2011) Pectoral fin loss in the Mastacembelidae: a new species from Lake Tanganyika. *Journal of Zoology*, **284**(4), 286–293

Brown, K. J., Rüber, L., Bills, R., & Day, J. J. (2010) Mastacembelid eels support Lake Tanganyika as an evolutionary hotspot of diversification. *BMC Evolutionary Biology*, **10**, 188

Chakrabarty, P. (2005) Testing conjectures about morphological diversity in cichlids of Lakes Malawi and Tanganyika. *Copeia*, **2005**(2), 359–373

Cohen, A. S., Soreghan, M. J., & Scholz, C. A. (1993a) Estimating the age of formation of lakes: An example from Lake Tanganyika , East African Rift system. *Geology*, **21**, 511–514

Cohen, A. S., Bills, R., Cocquyt, C. Z., & Caljon, A. G. (1993b) The impact of sediment pollution on biodiversity in Lake Tanganyika. *Conservation Biology*, **7**(3), 667–677

Cohen, A. S., Lezzar, K. E., Tiercelin, J. J., & Soreghan, M. (1997) New palaeogeographic and lake-level reconstructions of Lake Tanganyika: implications for tectonic, climatic and biological evolution in a rift lake. *Basin Research*, **9**(2), 107–132

Colombo, M., Diepeveen, E. T., Muschick, M., Santos, M. E., Indermaur, A., Boileau, N., & Salzburger, W. (2013) The ecological and genetic basis of convergent thick-lipped phenotypes in cichlid fishes. *Molecular Ecology*, **22**(3), 670–684

Coulter, G. W. (1991) *Lake Tanganyika and its Life*. Oxford: Oxford University Press.

Coulter, G. W., & Spigel, R. H. (1991) Hydrodynamics, In G. W. Coulter (Ed.), *Lake Tanganyika and its Life.* Oxford: Oxford University Press. pp. 49–75

Danley, P. D., & Kocher, T. D. (2001) Speciation in rapidly diverging systems: lessons from Lake Malawi. *Molecular Ecology*, **10**(5), 1075–1086

Day, J. J., Bills, R., & Friel, J. P. (2009) Lacustrine radiations in African Synodontis catfish. *Journal of Evolutionary Biology*, **22**(4), 805–817

Day, J. J., Cotton, J. A., & Barraclough, T. G. (2008) Tempo and mode of diversification of lake Tanganyika cichlid fishes. *PloS One*, **3**(3), e1730

Day, J. J., Peart, C. R., Brown, K. J., Friel, J. P., Bills, R., & Moritz, T. (2013) Continental diversification of an African catfish radiation (Mochokidae: Synodontis) *Systematic Biology*, **62**(3), 351-365

Day, J. J., Santini, S., & Garcia-Moreno, J. (2007) Phylogenetic relationships of the Lake Tanganyika cichlid tribe Lamprologini: the story from mitochondrial DNA. *Molecular Phylogenetics and Evolution*, **45**(2), 629–642

Day, J. J., & Wilkinson, M. (2006) On the origin of the Synodontis catfish species flock from Lake Tanganyika. *Biology Letters*, **2**(4), 548–552

Drummond, A. J., Ho, S. Y. W., Phillips, M. J., & Rambaut, A. (2006) Relaxed phylogenetics and dating with confidence. *PLoS Biology*, **4**(5), e88

Drummond, A. J., & Suchard, M. A. (2010) Bayesian random local clocks, or one rate to rule them all. *BMC Biology*, **8**, 114

Egger, B., Koblmüller, S., Sturmbauer, C., & Sefc, K. M. (2007) Nuclear and mitochondrial data reveal different evolutionary processes in the Lake Tanganyika cichlid genus Tropheus. *BMC Evolutionary Biology*, **7**, 137

Elmer, K. R., Fan, S., Kusche, H., Spreitzer, M. L., Kautt, A. F., Franchini, P., & Meyer, A. (2014) Parallel evolution of Nicaraguan crater lake cichlid fishes via non-parallel routes. *Nature Communications*, **5**, 5168

Erpenbeck, D., Weier, T., de Voogd, N. J., Wörheide, G., Sutcliffe, P., Todd, J. A, & Michel, E. (2011) Insights into the evolution of freshwater sponges (Porifera: Demospongiae: Spongillina): Barcoding and phylogenetic data from Lake Tanganyika endemics indicate multiple invasions and unsettle existing taxonomy. *Molecular Phylogenetics and Evolution*, **61**(1), 231–236

Friedman, M., Keck, B. P., Dornburg, A., Eytan, R. I., Martin, C. H., Darrin, C., & Near, T. J. (2013) Molecular and fossil evidence place the origin of cichlid fishes long after Gondwanan rifting Molecular and fossil evidence place the origin of cichlid fishes long after Gondwanan rifting. *Proceedings of the Royal Society B: Biological Sciences*, **280**(1770) 20131733

Fryer, G. (2006) Evolution in ancient lakes: radiation of Tanganyikan atyid prawns and speciation of pelagic cichlid fishes in Lake Malawi. *Hydrobiologia*, **568**(S1), 131–142

Fryer, G., & Iles, T. D. (1972) *The cichlid fishes of the Great Lakes of Africa*. Edinburgh: Oliver and Boyd

Fulton, M., & Hodges, S. A. (1999) Floral isolation between Aquilegia formosa and Aquilegia pubescens. *Proceedings of the Royal Society B: Biological Sciences*, **266**(1435), 2247–2252

Genner, M. J., Seehausen, O., Lunt, D. H., Joyce, D. A., Shaw, P. W., Carvalho, G. R., & Turner, G. F. (2007) Age of Cichlids: new dates for ancient lake fish radiations. *Molecular Biology and Evolution*, **24**(5), 1269–1282

Gonzalez-Voyer, A., Fitzpatrick, J. L., & Kolm, N. (2008) Sexual selection determines parental care patterns in cichlid fishes. *Evolution; International Journal of Organic Evolution*, **62**(8), 2015–2026

Grant, P. R. (1986) *Ecology and evolution of Darwin's finches*. Princeton: Princeton University Press

Hardman, M. (2008) A new species of catfish genus Chrysichthys from Lake Tanganyika (Siluriformes: Claroteidae) *Copeia*, **2008**(1), 43

Harmon, L. J., Weir, J. T., Brock, C. D., Glor, R. E., & Challenger, W. (2008) GEIGER: investigating evolutionary radiations. *Bioinformatics*, **24**(1), 129–131

Heled, J., & Drummond, A. J. (2010) Bayesian inference of species trees from multilocus data. *Molecular Biology and Evolution*, **27**(3), 570–580

Herder, F., Nolte, A. W., Pfaender, J., Schwarzer, J., Hadiaty, R. K., & Schliewen, U. K. (2006) Adaptive radiation and hybridization in Wallace's Dreamponds: evidence from sailfin silversides in the Malili Lakes of Sulawesi. *Proceedings of the Royal Society B: Biological Sciences,* **273**(1598), 2209–2217

Joyce, D., Lunt, D., Bills, R., Turner, G., Katongo, C., Duftner, N., & Seehausen, O. (2005) An extant cichlid fish radiation emerged in an extinct Pleistocene lake. *Nature*, **435**, 90-95

Katongo, C., Koblmüller, S., Duftner, N., Mumba, L., & Sturmbauer, C. (2007) Evolutionary history and biogeographic affinities of the serranochromine cichlids in Zambian rivers. *Molecular Phylogenetics and Evolution*, **45**(1), 326–338

Klerkx, J., Theunissen, K., & Delvaux, D. (1998) Persistent fault controlled basin formation since the Proterozoic along the Western Branch of the East African Rift. *Journal of African Earth Sciences*, **26**(3), 347–361

Koblmüller, S., Duftner, N., Sefc, K. M., Aibara, M., Stipacek, M., Blanc, M., & Sturmbauer, C. (2007) Reticulate phylogeny of gastropod-shell-breeding cichlids from Lake Tanganyika- the result of repeated introgressive hybridization. *BMC Evolutionary Biology*, **7**, 7

Koblmüller, S., Odhiambo, E. A., Sinyinza, D., Sturmbauer, C., & Sefc, K. M. (2014) Big fish, little divergence: phylogeography of Lake Tanganyika's giant cichlid, Boulengerochromis microlepis. *Hydrobiologia*. doi:10.1007/s10750-014-1863-z

Koblmüller, S., Schliewen, U. K., Duftner, N., Sefc, K. M., Katongo, C., & Sturmbauer, C. (2008) Age and spread of the haplochromine cichlid fishes in Africa. *Molecular Phylogenetics and Evolution*, **49**(1), 153–169

Koblmüller, S., Sturmbauer, C., Verheyen, E., Meyer, A., & Salzburger, W. (2006) Mitochondrial phylogeny and phylogeography of East African squeaker catfishes (Siluriformes: Synodontis) *BMC Evolutionary Biology*, **6**, 49

Kocher, T., Conroy, J., McKaye, K., & Stauffer, J. (1993) Similar morphologies of cichlid fish in lakes Tanganyika and Malawi are due to covergence. *Molecular Phylogenetics and Evolution*, **2**(2), 158–165

Lezzar, K. E., Tiercelin, J.-J., De Batist, M., Cohen, A. S., Bandora, T., Van Rensbergen, P., & Klerkx, J. (1996) New seismic stratigraphy and Late Tertiary history of the North Tanganyika Basin, East African Rift system, deduced from multichannel and high-resolution reflection seismic data and piston core evidence. *Basin Research*, **8**, 1–28

Liem, K. (1980) Adaptive significance of intra-and interspecific differences in the feeding repertoires of cichlid fishes. *American Zoologist*, **20**(1), 295–314

Liem, K. F. (1973) evolutionary strategies and morphological innovations: cichlid pharyngeal jaws. *Systematic Zoology*, **22**(4), 425

Losos, J. B. (2009) *Lizards in an evolutionary tree: ecology and adaptive radiation of anoles*. Oakland: University of California Press

Losos, J. B., Jackman, T. R., Larson, A., de Queiroz, K., & Rodriguez-Schettino, L. (1998) Contingency and determinism in replicated adaptive radiations of island lizards. *Science*, **279**(5359), 2115–2118

Mahler, D. L., Ingram, T., Revell, L. J., & Losos, J. B. (2013) Exceptional convergence on the macroevolutionary landscape in island lizard radiations. *Science*, **341**(6143), 292–295

Manousaki, T., Hull, P. M., Kusche, H., Machado-Schiaffino, G., Franchini, P., Harrod, C., & Meyer, A. (2013) Parsing parallel evolution: ecological divergence and differential gene expression in the adaptive radiations of thick-lipped Midas cichlid fishes from Nicaragua. *Molecular Ecology*, **22**(3), 650–669

Marijnissen, S. A. E., Michel, E., Daniels, S. R., Erpenbeck, D., Menken, S. B. J., & Schram, F. R. (2006) Molecular evidence for recent divergence of Lake Tanganyika endemic crabs (Decapoda: Platythelphusidae) *Molecular Phylogenetics and Evolution*, **40**(2), 628–634

Marijnissen, S. A. E., Michel, E., Kamermans, M., Olaya-Bosch, K., Kars, M., Cleary, D. F. R., & Menken, S. B. J. (2008) Ecological correlates of species differences in the Lake Tanganyika crab radiation. *Hydrobiologia*, **615**(1), 81–94

McGlue, M. M., Lezzar, K. E., Cohen, A. S., Russell, J. M., Tiercelin, J. J., Felton, A. A., & Nkotagu, H. H. (2008) Seismic records of late Pleistocene aridity in Lake Tanganyika, tropical East Africa. *Journal of Paleolimnology*, **40**(2), 635–653

McIntyre, P. B., Jones, L. E., Flecker, A. S., & Vanni, M. J. (2007) Fish extinctions alter nutrient recycling in tropical freshwaters. *Proceedings of the National Academy of Sciences*, **104**(11), 4461–4466

Meyer, A., Knowles, L. L., & Verheyen, E. (1996) Widespread geographical distribution of mitochondrial haplotypes in rock-dwelling cichlid fishes from Lake Tanganyika. *Molecular Ecology*, **5**(3), 341–350

Mo, T. P. (1991) *Theses Zoologicae vol. 17: Anatomy, relationships and systematics of the Bagridae (Teleostei: Siluroidei)—with a hypothesis of siluroid phylogeny*. Koenigstein: Koeltz Scientific Books

Muschick, M., Indermaur, A., & Salzburger, W. (2012) Convergent evolution within an adaptive radiation of cichlid fishes. *Current Biology*, **22**(24), 2362–2368

O'Reilly, C. M., Alin, S. R., Plisnier, P. D., Cohen, A. S., & McKee, B. A. (2003) Climate change decreases aquatic ecosystem productivity of Lake Tanganyika, Africa. *Nature*, **424**(6950), 766–768

Ochi, H., Rossiter, A., & Yanagisawa, Y. (2000) The first record of a biparental mouthbrooding catfish. *Journal of Fish Biology*, **57**(6), 1601–1604

Ochi, H., Rossiter, A., & Yanagisawa, Y. (2001) Biparental mouthbrooding of the catfish Phyllonemus filinemus in Lake Tanganyika. *Ichthyological Research*, **48**, 225–229

Ochi, H., Rossiter, A., & Yanagisawa, Y. (2002) Paternal mouthbrooding bagrid catfishes in Lake Tanganyika. *Ichthyological Research*, **49**, 270–273

Revell, L. J. (2012) phytools: an R package for phylogenetic comparative biology (and other things) *Methods in Ecology and Evolution*, **3**(2), 217–223

Salzburger, W. (2009) The interaction of sexually and naturally selected traits in the adaptive radiations of cichlid fishes. *Molecular Ecology*, **18**(2), 169–185

Salzburger, W., Baric, S., & Sturmbauer, C. (2002) Speciation via introgressive hybridization in East African cichlids? *Molecular Ecology*, **11**(3), 619–625

Salzburger, W., Mack, T., Verheyen, E., & Meyer, A. (2005) Out of Tanganyika: genesis, explosive speciation, key-innovations and phylogeography of the haplochromine cichlid fishes. *BMC Evolutionary Biology*, **5**(1983), 17

Sander, P. M., Klein, N., Buffetaut, E., Cuny, G., Suteethorn, V., & Le Loeuff, J. (2004) Adaptive radiation in sauropod dinosaurs: bone histology indicates rapid evolution of giant body size through acceleration. *Organisms, Diversity & Evolution*, **4**, 165–173

Sato, T. (1986) A brood parasitic catfish of mouthbrooding cichlid fishes in Lake Tanganyika. *Nature*, **323**, 58–59

Schliewen, U. K., Tautz, D., & Paabo, S. (1994) Sympatric speciation suggested by monophyly of crater lake cichlids. *Nature*, **368**, 629–632

Schluter, D. (2000) *The Ecology of Adaptive Radiation*. Oxford: Oxford University Press

Schluter, D., & Conte, G. L. (2009) Genetics and ecological speciation. *Proceedings of the National Academy of Sciences*, **106**, 9955–9962

Scholz, C. A., King, J. W., Ellis, G. S., Swart, P. K., Stager, J. C., & Colman, S. M. (2003) Paleolimnology of Lake Tanganyika , East Africa, over the past 100 k yr. *Journal of Paleolimnology*, **30**, 139–150

Schön, I., & Martens, K. (2011) Molecular analyses of ostracod flocks from Lake Baikal and Lake Tanganyika. *Hydrobiologia*, **682**(1), 91–110

Seehausen, O. (2006) African cichlid fish: a model system in adaptive radiation research. *Proceedings of the Royal Society B: Biological Sciences*, **273**(1597), 1987–1998

Sefc, K. M., Baric, S., Salzburger, W., & Sturmbauer, C. (2007) Species-specific population structure in rock-specialized sympatric cichlid species in Lake Tanganyika, East Africa. *Journal of Molecular Evolution*, **64**(1), 33–49

Simpson, G. G. (1953) *The major features of evolution*. New York: Columbia University Press

Sturmbauer, C., Mark, W., & Dallinger, R. (1992) Ecophysiology of Aufwuchs-eating cichlids in Lake Tanganyika: niche separation by trophic specialization. *Environmental Biology of Fishes*, **35**, 283–290

Sturmbauer, C., Salzburger, W., Duftner, N., Schelly, R., & Koblmüller, S. (2010) Evolutionary history of the Lake Tanganyika cichlid tribe Lamprologini (Teleostei: Perciformes) derived from mitochondrial and nuclear DNA data. *Molecular Phylogenetics and Evolution*, **57**(1), 266–284

Sugawara, T., Terai, Y., Imai, H., Turner, G. F., Koblmüller, S., Sturmbauer, C., & Okada, N. (2005) Parallelism of amino acid changes at the RH1 affecting spectral sensitivity among deep-water cichlids from Lakes Tanganyika and Malawi. *Proceedings of the National Academy of Sciences*, **102**(15), 5448–5453

Sullivan, J. P., Lundberg, J. G., & Hardman, M. (2006) A phylogenetic analysis of the major groups of catfishes (Teleostei: Siluriformes) using rag1 and rag2 nuclear gene sequences. *Molecular Phylogenetics and Evolution*, **41**(3), 636–662

Tierney, J., Mayes, M., & Meyer, N. (2010) Late-twentieth-century warming in Lake Tanganyika unprecedented since AD 500. *Nature Geoscience*, **3**, 422–425

Van Steenberge, M., Vanhove, M. P. M., Muzumani Risasi, D., Mulimbwa N'Sibula, T., Muterezi Bukinga, F., Pariselle, A., & Snoeks, J. (2011) a recent inventory of the fishes of the north-western and central western coast of Lake Tanganyika (Democratic Republic Congo) *Acta Ichthyologica Et Piscatoria*, **41**(3), 201–214

Van Zwieten, P. A. M., Roest, F. C., Machiels, M. A. M., & van Densen, W. L. T. (2002) Effects of inter-annual variability, seasonality and persistence on the perception of long-term trends in catch rates of the industrial pelagic purse-seine fishery of northern Lake Tanganyika (Burundi) *Fisheries Research*, **54**, 329–348

Verburg, P., & Hecky, R. (2009) The physics of the warming of Lake Tanganyika by climate change. *Limnology and Oceanography*, **54**(6), 2418–2430

Wagner, C. E., Harmon, L. J., & Seehausen, O. (2012a) Ecological opportunity and sexual selection together predict adaptive radiation. *Nature*, **487**(7407), 366–369

Wagner, C. E., Keller, I., Wittwer, S., Selz, O. M., Mwaiko, S., Greuter, L., & Seehausen, O. (2013) Genome-wide RAD sequence data provide unprecedented resolution of species boundaries and relationships in the Lake Victoria cichlid adaptive radiation. *Molecular Ecology*, **22**(3), 787–798

Wagner, C. E., & McCune, A. R. (2009) Contrasting patterns of spatial genetic structure in sympatric rock-dwelling cichlid fishes. *Evolution*, **63**(5), 1312–1326

Wagner, C. E., McCune, A. R., & Lovette, I. J. (2012b) Recent speciation between sympatric Tanganyikan cichlid colour morphs. *Molecular Ecology*, **21**(13), 3283–3292

West, K., & Michel, E. (2000) The dynamics of endemic diversification: Molecular phylogeny suggests an explosive origin of the thiarid gastropods of Lake Tanganyika. In Rossiter, A., & Kawanabe, H. (Eds.) *Advances in Ecological Research* **31**. pp. 331–354

Wilson, A. B., Glaubrecht, M., & Meyer, A. (2004) Ancient lakes as evolutionary reservoirs: evidence from the thalassoid gastropods of Lake Tanganyika. *Proceedings of the Royal Society B: Biological Sciences*, **271**(1538), 529–536

Wilson, G. P., Evans, A. R., Corfe, I. J., Smits, P. D., Fortelius, M., & Jernvall, J. (2012) Adaptive radiation of multituberculate mammals before the extinction of dinosaurs. *Nature*, **483**(7390), 457–460

Winkelmann, K., Genner, M. J., Takahashi, T., & Rüber, L. (2014) Competition-driven speciation in cichlid fish. *Nature Communications*, **5**, 3412

Wright, J. J. (2011) Conservative Coevolution of Müllerian Mimicry in a Group of Rift Lake Catfish. *Evolution*, **65**(2), 395–407

Wright, J. J., & Bailey, R. M. (2012) Systematic revision of the formerly monotypic genus Tanganikallabes (Siluriformes: Clariidae) *Zoological Journal of the Linnean Society*, **165**(1), 121–142

Wright, J. J., & Page, L. M. (2006) Taxonomic revision of Lake Tanganyikan Synodontis (Siluriformes: Mochokidae) *Bulletin of the Florida Museum of Natural History*, **46**(4), 99–154

Zoppoth, P., Koblmüller, S., & Sefc, K. M. (2013) Male courtship preferences demonstrate discrimination against allopatric colour morphs in a cichlid fish. *Journal of Evolutionary Biology*, **26**(3), 577–586

Chapter 2
Nocturnal claroteine catfishes reveal dual colonisation but a single radiation in Lake Tanganyika

2.1 Abstract

Lake Tanganyika (LT) is a biodiversity hotspot supporting many endemic radiations that provide comparative systems in which to investigate if there are common factors leading to the build-up of its considerable diversity. Despite LT containing the highest diversity of lacustrine catfishes on Earth, the evolutionary relationships of nocturnal catfishes within the sub-family Claroteinae have not been investigated and it is unknown if its constituent genera have diversified via single or independent colonisation events. In this study, the first molecular phylogeny of the LT claroteine catfishes is presented, based on a multigene dataset (three nuclear markers, two mitochondrial totalling 4227 bp), including 85 samples from LT and outside of the lake basin. These data support LT claroteine monophyly, with the exclusion of the LT endemic *Chrysichthys brachynema* that independently colonised the lake but has not radiated. Multiple sampling localities from LT and the use of Bayesian species delimitation methods reveal additional locally restricted diversity within the LT Claroteinae clade. Fossil calibrated molecular divergence dates suggest that diversification occurred within full lake conditions as demonstrated in other LT lineages.

2.2 Introduction

Insular faunas are widely considered to be particularly well suited to studies assessing which factors generate and maintain diversity (e.g., MacArthur and Wilson, 1967; Wallace, 1880). The great lakes of East Africa's Rift Valley represent a system of freshwater 'islands' and as a result have attracted much attention, not least because of their hyper-diverse, endemic cichlid fish radiations and interest in the possible mechanisms responsible for generating this spectacular diversity (e.g., Day et al., 2008; Genner and Turner, 2012; Haesler and Seehausen, 2005; Muschick et al., 2012). Additionally, Lake Malawi and Lake Tanganyika host a number of independent radiations facilitating the search for general patterns and causes.

One approach to identifying common patterns across multiple radiations is to look at the same or closely related taxa in different environments. For example, Wagner et al. (2012) investigated cichlid colonisation and diversification in lakes throughout Africa and found that deep lakes, high net solar radiation, more time for diversification and the presence of sexual selection are associated with diversification but lake surface area is not. This suggests that radiations may be predictable to some extent in terms of propensity to diversify if not necessarily in terms of exact outcomes. For example, whilst repeated phenotypes have evolved in different East African Rift Lakes (Fryer and Iles, 1972), smaller cichlid radiations in crater lakes (≤11 species) that are similar in size to other non-cichlid Rift Lake radiations, do not show equivalent species assemblages (Schliewen et al., 1994).

An alternative way to determine if there are common factors facilitating the generation of biological diversity is to investigate the similarities and differences between radiations in divergent taxa within the same environment. Among the East African Rift Lakes, Lake Tanganyika (LT) offers a uniquely broad diversity of different-sized endemic radiations from independent lineages with differing ecologies and life histories e.g., gastropods (approximately 70 species, West and Michel, 2000), copepods (68 species and sub-species, Coulter, 1991), *Mastacembelus* eels (14 species, Brown et al., 2010, 2011), *Synodontis* catfishes (~11 species, Day and Wilkinson, 2006; Day et al., 2009; Wright and Page, 2006), *Platythelphusa* crabs (nine species, Marijnissen et al., 2006), *Tanganikallabes* catfishes (three species, Wright and Bailey, 2012). Cichlid fishes may not be representative of other taxa due to their exceptional ability to speciate. Thus, comparative data from other radiations are important in testing whether patterns of diversification in cichlids are also seen across a range of other lineages.

A further advantage of LT as a study system is its long history as a water body and relative stability allowing increased time for colonisation and diversification. Lake Tanganyika is the oldest of the East African Lakes at 9–12 Ma (Cohen et al., 1993) and has experienced periods of low lake levels that led to the separation of different basins (Cohen et al., 1997) but has not fully dried out.

In order to compare radiations, age estimates for each clade are required. Recent methodological advances in the construction of species trees using multispecies coalescence offer the possibility of improved estimation of speciation times for recently diverged taxa taking into account genetic divergence prior to

speciation (Drummond et al., 2012; McCormack et al., 2011). In addition molecular species delimitation offers an objective way to define the evolutionarily significant units used in analyses based on multiple loci ensuring that equivalent taxonomic units are compared, which will also facilitate future comparison with other radiations.

2.2.1 The study system

Catfishes of the sub-family Claroteinae from LT offer a useful system to study lacustrine diversification due to the inclusion of multiple genera, encompassing widespread and locally restricted species that occur at different depths (Hardman, 2008), as well as rocky shore and deep water specialists. Four claroteine genera have representatives in LT. Three genera are endemic: the monotypic genus *Bathybagrus* (Bailey and Stewart, 1984), *Lophiobagrus* (four species; Bailey and Stewart, 1984) and *Phyllonemus*. Within *Phyllonemus* there are three currently described species, (Risch, 1987), although morphological variation is suggestive of additional species (Bills, pers. obs.). The fourth genus, *Chrysichthys*, is widespread throughout Africa, with seven species currently described in LT (Hardman, 2008), and a further 40 elsewhere (Eschmeyer, 2013). There is some controversy over the taxonomy of the group with Mo (1991) expanding the genus *Bathybagrus* and reassigning some species from *Chrysichthys*. However, Mo's taxonomy is disputed, (Hardman, 2008), and given that the inconsistencies in Mo's study have not been resolved the taxonomy of Bailey and Stewart (1984) is maintained in this study.

The molecular phylogeny presented in this study is the first for the LT claroteines and incorporates sampling from both the northern and southern basins. In addition it includes morphologically divergent specimens putatively of the same species. This is important, because species diversity in LT catfishes has previously been underestimated, as evidenced by recent increases in described species, including several known from a single locality (Hardman, 2008; Wright and Bailey, 2012; Wright and Page, 2006).

Most claroteine catfishes within LT inhabit the littoral zone, however several *Chrysichthys* species and *Bathybagrus tetranema* are found in deeper water, occupying a similar niche to the only other endemic lacustrine catfish radiation outside of LT, the deep water *Clarias-Bathyclarias* radiation of Lake Malawi (Agnèse and Teugels, 2001). Major habitat shifts have been identified as

the first stage of adaptive radiation in Lake Malawi cichlids (reviewed in Danley and Kocher, 2001) and account for initial divergence and early acceleration of diversification in several marine adaptive radiations e.g., seven-spined gobies (Rüber et al., 2003) and scarids (Streelman et al., 2002).

Here, the evolutionary history, diversification and systematics of LT claroteine catfishes are investigated with a dated multi-gene molecular phylogeny, in order to answer the following questions; i) Were there multiple independent colonisations or a single event? ii) Did diversification occur within LT (a source of diversity) or does it contain multiple ancient lineages (a biodiversity sink)? iii) Did the different genera colonise and diversify simultaneously, indicating a common external driver of diversification? Additionally, whether the overall diversity of claroteine catfishes is currently underestimated is tested by using sampling from around the lake and employing Bayesian species delimitation to assess if the observed morphological diversity in *Phyllonemus* corresponds to increased species diversity. Finally, the phylogeny is used to determine if the traditional taxonomy of the group is supported, or whether there is any support for the alternative taxonomy of Mo (1991).

2.3 Methods

2.3.1 Sampling

The specimens included in this study were collected using a variety of methods e.g., scuba, snorkelling, fyke nets, rotenone and obtained in local fish markets. The majority of specimens are preserved as vouchers; a full list of the taxa in this study is given in Appendix 1 (Table 1). Samples were collected from both the northern and southern basins within LT from Burundi, Tanzania and Zambia. Complete sampling of LT is, however, problematic in light of the large portion of LT in the Democratic Republic of the Congo and its ongoing political instability.

In total 85 specimens are included in this study, representing 10 of the 15 described LT claroteine species (and additional specimens not identified to species level) and four putative new *Phyllonemus* species. An additional 21 claroteine specimens from outside LT were included expanding the sampling to include 5 of the 8 claroteine genera. The following claroteine genera are not included in this study: *Amarginops* (single species), *Gephyroglanis* (three species) and *Pardiglanis*

(single species). Many *Chrysichthys* specimens in museum collections are not identified to species level because they are often difficult to assign based on morphological features, so samples were chosen to maximise geographic coverage in an attempt to maximise genetic diversity. The species *Rheoglanis dendrophorus* is not included here, although it has been used in a previous study dating divergences involving the Claroteidae (Lundberg et al., 2007). There is continuing taxonomic uncertainty regarding the validity of this genus and at present *Rheoglanis* has been synonymised with *Chrysichthys* (Eschmeyer, 2013).

Outgroup taxa (eight samples, Appendix 1, Table 1) were selected from the Auchenoglaninae (*Auchenoglanis occidentalis*, *Parauchenoglanis ngamensis* and *Parauchenoglanis monkei*), the other sub-family of the Claroteidae (Mo, 1991), and the family Schilbeidae (*Parailia congica*, *Pareutropius debauwi* and *Schilbe intermedius*).

2.3.2 DNA extraction, amplification and sequencing

A combination of mitochondrial and nuclear genes was sequenced (Appendix 1, Table 1); the mitochondrial genes Cytochrome c oxidase sub-unit 1 (CO1 652 bp) and Cytochrome b (*Cytb* 1138 bp) provide shallow phylogenetic resolution, CO1 is the barcoding gene and *Cytb* has been shown to provide good resolution in the *Synodontis* catfish radiation within LT (Day and Wilkinson, 2006). Three nuclear genes were sequenced: the ribosomal coding gene S7 intron 1 (S7 724 bp) provided resolution in *Synodontis* (Day et al., 2009), Recombination Activating Gene 2 (RAG2 912 bp) has been used in a variety of catfish phylogenetic studies (e.g., Day et al., 2013; Hardman and Page, 2003; Sullivan et al., 2006) and Plagl2 (801 bp) is a single copy nuclear gene, identified by genome comparison, shown to provide deeper resolution in teleosts (Li et al., 2007).

DNA was extracted using a Qiagen DNeasy Blood and Tissue Extraction Kit (Qiagen Ltd). Primers and PCR conditions were used from the following studies: CO1 (Folmer et al., 1994; Ward et al., 2005), *Cytb* (Brown et al., 2010; Hardman, 2005; Hardman and Page, 2003), RAG2 (Hardman and Page, 2003), Plagl2 (Li et al., 2007) and S7 (Chow and Hazama, 1998). The novel *Cytb* primer I*Cytb*: CCTACATGAAACRGGCTCAA was also used in conjunction with FishProR1 (Brown et al., 2010) and the following PCR conditions: 94°C for 3 min, followed by (94°C for 30 s annealing temperature 60°C for 30 s, 72°C for 45 s) × 35 followed by 72°C

for 5 min. The protein coding genes (CO1, *Cytb*, RAG2 and Plagl2) were aligned in Geneious 5.6 (Biomatters) using default settings with alignments checked for stop codons and reading frame shifts. The S7 intron was aligned using Geneious consensus align (70% similarity matrix, gap open penalty 17, gap extension penalty 3). There were few instances of heterozygous positions in the nuclear genes and these were coded as ambiguous data using IUPAC ambiguity codes.

2.3.3 Phylogenetic inference

The optimal partitioning scheme and model choices were assessed with PartitionFinder 1.0 (Lanfear et al., 2012) using the greedy algorithm, considering only models available in MrBayes (Huelsenbeck and Ronquist, 2001) and assessed using Bayesian Information Criteria (BIC) with 13 possible subsets defined (each codon position for the protein coding genes plus S7). The concatenated mitochondrial and concatenated nuclear datasets were also analysed separately.

Phylogenetic analyses were conducted in MrBayes 3.1.2 (Huelsenbeck and Ronquist, 2001). Each analysis was conducted with two separate runs starting from random trees for ten million generations (15 million for the concatenated analysis) using four Markov chains (three heated, one cold, heating parameter 0.1) with default priors. Burn-in was assessed graphically in Tracer 1.5 (Rambaut and Drummond, 2009) and the results from both independent runs were combined. Stationarity of the chains was also investigated in Tracer 1.5 by plotting the likelihood score and all parameter values against the number of generations to ensure that they had stabilised. In an attempt to avoid stabilising at local maxima each analysis was repeated twice and convergence was assessed graphically using AWTY (Nylander et al., 2008). In the event that chains did not reach convergence or if the rate of reaching convergence was slow then the heating parameter was halved. This led to the concatenated nuclear dataset having a heating parameter of 0.05 and the complete concatenated dataset having a heating parameter of 0.025.

2.3.4 Molecular dating

Each nuclear gene and the concatenated mitochondrial dataset were investigated for saturation by plotting the uncorrected P distance against the models and parameters from Jmodeltest (Posada, 2008). The mitochondrial dataset was found to be saturated and therefore only the nuclear dataset was used for the dating

analysis as saturation can bias date estimates in fossil calibrated trees (e.g., Brandley et al., 2011; Lukoschek et al., 2012; Phillips, 2009).

The fossil record for extant claroteine genera consists of *Chrysichthys mahengeensis* (Murray and Budney, 2003) and more recent *Clarotes* specimens (5–7 Ma) that are not identified to species level (Stewart, 1994). The lake containing fossils described as *C. mahengeensis* (Murray and Budney, 2003) has been dated using Pb–U to 45.83 Ma ± 0.17 (Harrison et al., 2001). This species was used as a calibration point in a previous study investigating intercontinental siluriform relationships (Lundberg et al., 2007). However, this analysis reveals the non-monophyly of *Chrysichthys*, with the majority of LT taxa (those without a post-cleithral process) forming a well-supported clade that does not include the type species (*C. auratus*). *Chrysichthys mahengeensis* was described based on comparison to an unidentified *Chrysichthys* specimen (CMN 81-0240) from LT, which was examined in this study and determined not to have a post-cleithral process. Following the key in Hardman (2008) this specimen is tentatively identified as *C. platycephalus* and it is therefore not a member of the sub-clade of *Chrysichthys* that includes the type species. Until this fossil can be compared to a validated member of *Chrysichthys* it is not possible to accurately place it on a phylogeny and so it was not applied as a constraint in this analysis.

The oldest catfish fossil in Africa has been identified as a member of the extinct genus *Nigerium* from the Late Palaeocene (Landenian) (White, 1935) and was placed within the Claroteidae (Longbottom, 2010), however, this genus cannot be assigned to either the Claroteinae or the Auchenoglaninae. There is conflicting evidence as to whether these two sub-families are sister taxa. Sullivan et al., (2006) found that the Claroteinae were sister to the Schilbeidae based on 3660 bp from RAG1 and RAG2 sequences. Lundberg et al., (2007) extended the dataset to include the Mesoamerican catfish *Lacantunia enigmatica* (not included in the present study) and found that this taxon was sister to the Claroteinae with the Auchenoglaninae more closely related than the Schilbeidae. In spite of the taxonomic uncertainty of these sub-families, in the absence of other reliable fossils this calibration was used to date the phylogeny. Fossil calibration is preferred here to geological dates based on the age of LT, since these do not allow independent investigation of colonisation and diversification scenarios.

Molecular dating was implemented including only the claroteid taxa in BEAST 1.7.4 (Drummond et al., 2012). In order to account for uncertainty in the date estimate and to treat the fossil as a minimum age constraint log-normal priors were applied (mean = 0, stdev = 1, offset = 55) to the root. Preliminary analysis showed that yule and birth–death priors behaved similarly, so the simpler model implementing the yule prior was selected. RAG2 and Plagl2 were found to be evolving under a strict molecular clock whereas S7 was assigned a relaxed log-normal clock. The nuclear dataset was pruned to include one representative from each clade in the concatenated analysis in order to use a model of speciation with no coalescence whereas *BEAST used the complete dataset. The species assignations were based on previous taxonomic knowledge combined with the results of the MrBayes analyses in order to incorporate the cryptic diversity. Each putative *Phyllonemus* species was included separately. The optimal models and partitioning schemes for the reduced dataset were selected using PartitionFinder 1.0. Each analysis was completed three times, the BEAST analysis with a chain length of 5×10^7 (the influence of the priors was also investigated using an empty alignment) and the *BEAST analysis with a chain length of 25×10^7. Burn-in was determined graphically in Tracer 1.5 and convergence assessed as for the MrBayes analyses.

2.3.5 Species delimitation

This study encompasses several putative *Phyllonemus* species that differ morphologically (Bills pers. obs.). In order to assess the validity of these putative species a method of Bayesian species delimitation was implemented using the program Bayesian Phylogenetics and Phylogeography (BP&P 2.0) (Yang and Rannala, 2010). The analysis included 19 specimens representing seven putative (based on morphological observations) species (one sample of *Phyllonemus* sp. D, three samples of the other putative species) and used the three nuclear markers from the phylogenetic analysis (RAG2, Plagl2 and S7) with a total of four sequences missing in their entirety (Appendix 1, Table 1).

The program BP&P takes into account the species phylogeny and can accommodate incomplete lineage sorting. The age of the root of the species tree (s0) was assigned a gamma prior (25, 5000) and a mean mutation rate of 0.005 (estimated from the calibrated BEAST analysis). At present the putative

Phyllonemus species appear to be locally restricted and therefore likely to have small population sizes but there is no prior information on ancestral population sizes. Following the suggestion of Zhang et al., (2011) that mean values between 0.001 and 0.01 are representative of many species, the analysis was conducted once with the gamma prior 2, 2000, (mean = 0.001) and once with the gamma prior 1, 100 (mean = 0.01). This represents a conservative approach to species delimitation because with values larger than the true parameters simulation results suggest that the algorithm will be more likely to ‘‘lump’’ species together than to find artificial splits (Yang and Rannala, 2010). Ancestral population sizes can be estimated in *BEAST, however, the main factor in estimating population sizes is the number of independent loci (Heled and Drummond, 2010). Given that only three independent (nuclear) markers were available and the low number of sequences per species (three), using priors corresponding to a range of population sizes was preferable.

The topology of the guide tree can affect the results in a BP&P analysis (Leaché and Fujita, 2010). In this study the concatenated MrBayes phylogeny (nuclear and mitochondrial data) was selected as the guide tree because this provided the best supported phylogeny. However, it is notable that the mitochondrial MrBayes phylogeny and the nuclear BEAST phylogeny provide strong support for some alternative species placements within *Phyllonemus* (Figure 3; Figure 4). The extent of possible nuclear-mitochondrial discordance in this group is unknown based on the data presented here, representing a limitation to this study.

To account for the different mutation rates of the S7 intron compared to the protein coding genes, two approaches were considered: one with the rates drawn from a dirichlet prior with $\alpha = 2$, the other with the rates of evolution of each gene from the BEAST analysis used as an input variable. In order to assess the reliability of the results, each analysis was conducted using both species delimitation algorithms available in BP&P, each with a random starting seed. Each analysis was run for 500,000 generations with 10,000 generations discarded as burn-in.

2.4 Results

2.4.1 Model selection and phylogenetic analysis

The optimal partitioning scheme for the concatenated dataset was: Subset 1, 3rd codon positions of CO1 and *Cytb* (GTR + I + G), subset 2, 1st and 2nd positions of CO1 and *Cytb*, all codon positions of RAG2 and Plagl2, and S7 (SYM + I + G).

All analyses showed the non-monophyly of the LT claroteines (Figure 1). With the exception of the morphologically distinctive endemic LT basin taxon *Chrysichthys brachynema* all other LT claroteines were monophyletic and are hereafter referred to as the LT clade. In contrast *C. brachynema* was sister to a Zambian specimen of *Chrysichthys mabusi* within the clade containing the type species (*Chrysichthys auratus*) of the genus. Analysis of the concatenated dataset resolved the described genera within the LT clade as monophyletic with the exception of *Chrysichthys*, of which *C. sianenna* was sister to all other taxa in the LT clade (Figure 1). The monotypic genus *Bathybagrus* was sister to *Chrysichthys* within the LT clade (excluding *C. sianenna* and *C. brachynema*, hereafter referred to as LT *Chrysichthys*). The genus *Phyllonemus* was well supported, although intrarelationships varied between the mitochondrial and nuclear analyses. Within the well-supported genus *Lophiobagrus* (1 Bayesian Posterior Probability; BPP) the position of *L. brevispinis* as sister to a clade containing *L. cyclurus* and *L. aquilus* was well-supported (1 BPP). However, within this clade, there was previously unrecorded diversity with *L. cyclurus* resolving into three distinct groups, one containing specimens from Zambia, another containing specimens from Tanzania and a third containing specimens from both Tanzania and Burundi. The relationships between each of these *L. cyclurus* groups and *L. aquilus* were unresolved. *Clarotes laticeps* was resolved as sister to the LT clade rather than a species of *Chrysichthys* as previously proposed (Bailey and Stewart, 1984).

The nuclear tree was generated using a partitioning scheme with a single subset and the model SYM+I+G (Figure 2). A mitochondrial tree was generated using a partitioning scheme with two subsets: Subset 1, the 3rd positions of CO1 and *Cytb* (GTR + I + G), subset 2, 1st and 2nd positions of both CO1 and *Cytb* (GTR + I + G) (Figure 3). The mitochondrial tree, in contrast with the nuclear tree, resolved *Lophiobagrus* as sister to the other LT taxa with the exception of *C. brachynema*. This result showed conflicting signals between the nuclear and mitochondrial

datasets regarding the placement of *C. sianenna*, which may be due to lack of data or real conflict. The relationships between the different genera were not supported in the mitochondrial analysis but in both the nuclear and mitochondrial analyses the genera *Phyllonemus*, *Lophiobagrus* and LT *Chrysichthys* were well supported.

2.4.2 Estimation of molecular divergence dates

The empty BEAST alignment showed that the priors were similar in runs with and without data and demonstrated that the data were informative by producing different posteriors when data were added. Overall the *BEAST dates were more recent than those inferred with BEAST but were nonetheless broadly concordant (Figure 4), with the LT clade dated at 3.62–6.54 Ma 95% Highest Posterior Distribution (HPD) by BEAST and 2.88–5.95 Ma (HPD) in *BEAST. Within the principal LT radiation, the three genera diversified recently, with *Lophiobagrus* dated at 2.54–5.18 Ma (BEAST HPD), 1.97–4.61 Ma (*BEAST HPD), LT *Chrysichthys* dated at 0.77–2.67 Ma (BEAST HPD) and 0.43–1.76 Ma (*BEAST HPD) and *Phyllonemus* dated at 1.34–3.12 Ma (BEAST HPD) and 0.88– 2.52 Ma (*BEAST HPD).

The placement of *Bathybagrus tetranema* was not well supported in either the BEAST or the *BEAST dated phylogenies and as such this divergence cannot be dated. The BEAST and *BEAST analyses offered differing placements for *C. sianenna*, with this taxon strongly supported as sister to remaining taxa in the LT clade in the former analysis, whereas it was placed sister to a clade containing LT *Chrysichthys*, *Phyllonemus* and *Bathybagrus* in the latter analysis. This placement was not well supported in the *BEAST analysis, however the branch uniting the clade containing, *C. sianenna*, LT *Chrysichthys*, *Phyllonemus* and *Bathybagrus* to the exclusion of *Lophiobagrus* was well supported (0.99 BPP). The nuclear data alone did not resolve the relationships between the *Phyllonemus* species in the MrBayes analysis, however in the reduced BEAST analysis some branches in *Phyllonemus* were strongly supported and suggest an alternative *Phyllonemus* taxonomy to that found in the mitochondrial tree. In the *BEAST analysis the relationships between species in *Phyllonemus* were not well supported.

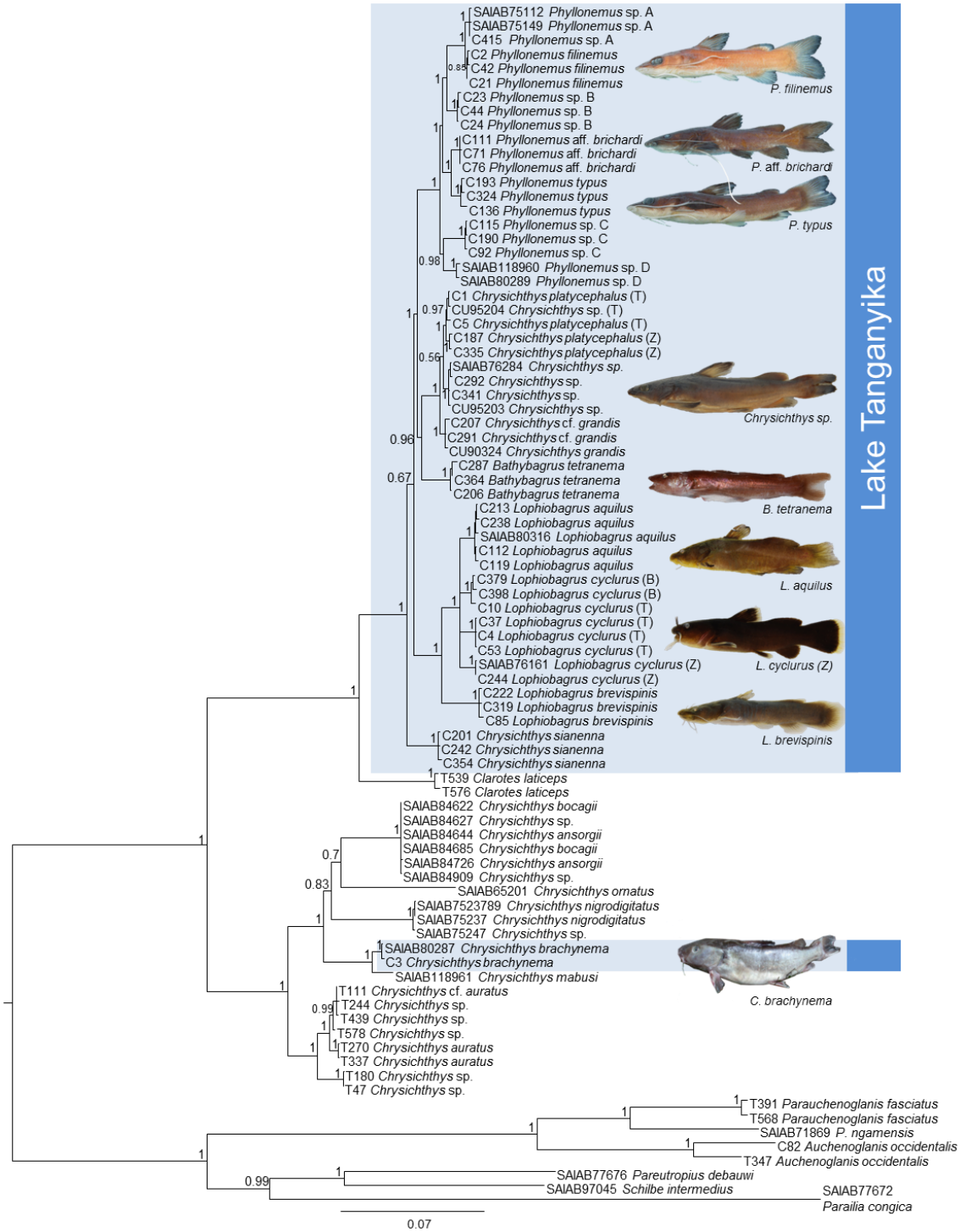


Figure 1 Consensus tree generated from the concatenated MrBayes analysis. Numbers at the nodes are Bayesian posterior probability (BPP) values. Taxa contained within the two boxes are those found in Lake Tanganyika. Photos R. Bills

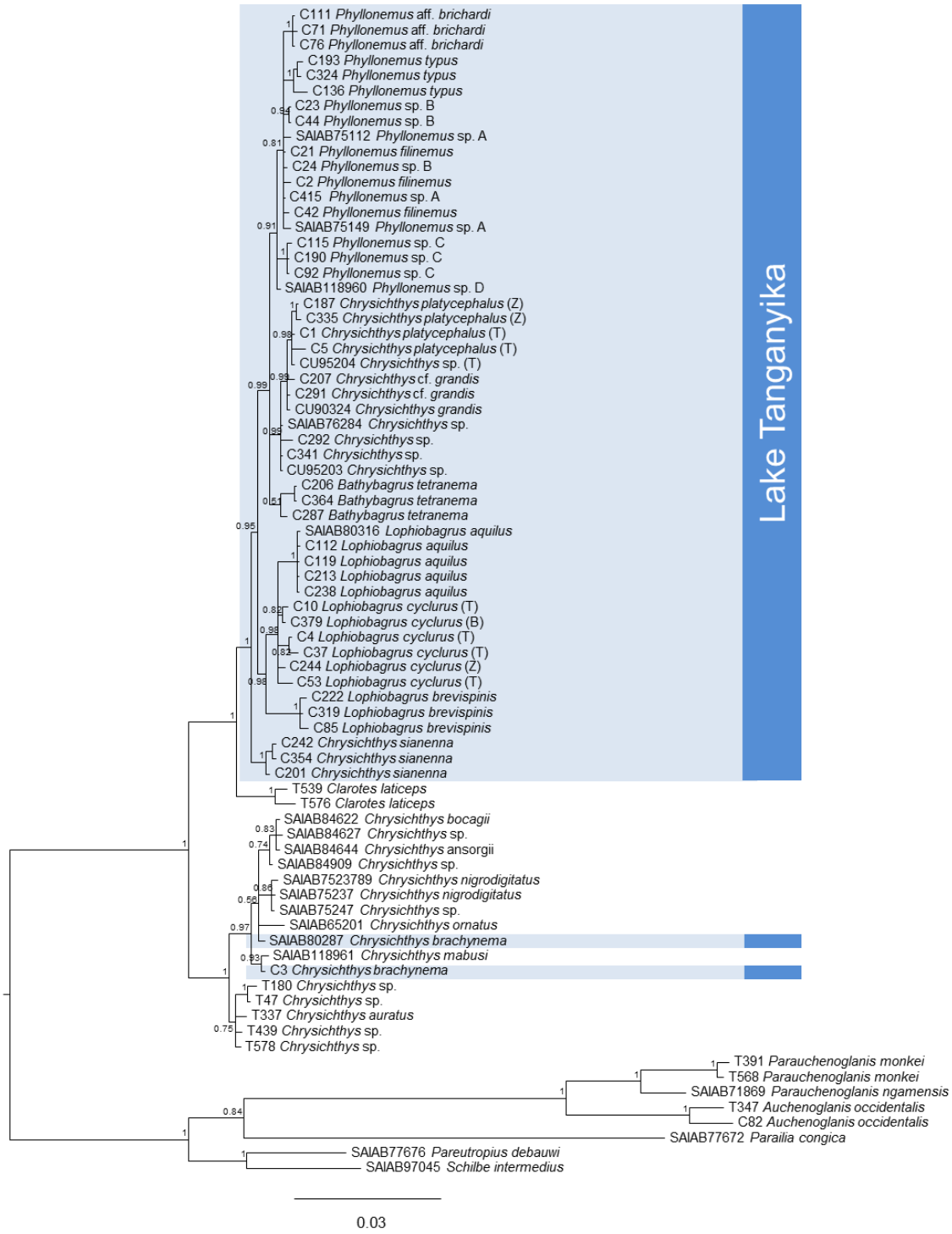


Figure 2 Consensus tree generated from the concatenated nuclear MrBayes analysis. Numbers at the nodes are Bayesian posterior probability (BPP) values. Taxa contained within the boxes are those found in Lake Tanganyika.

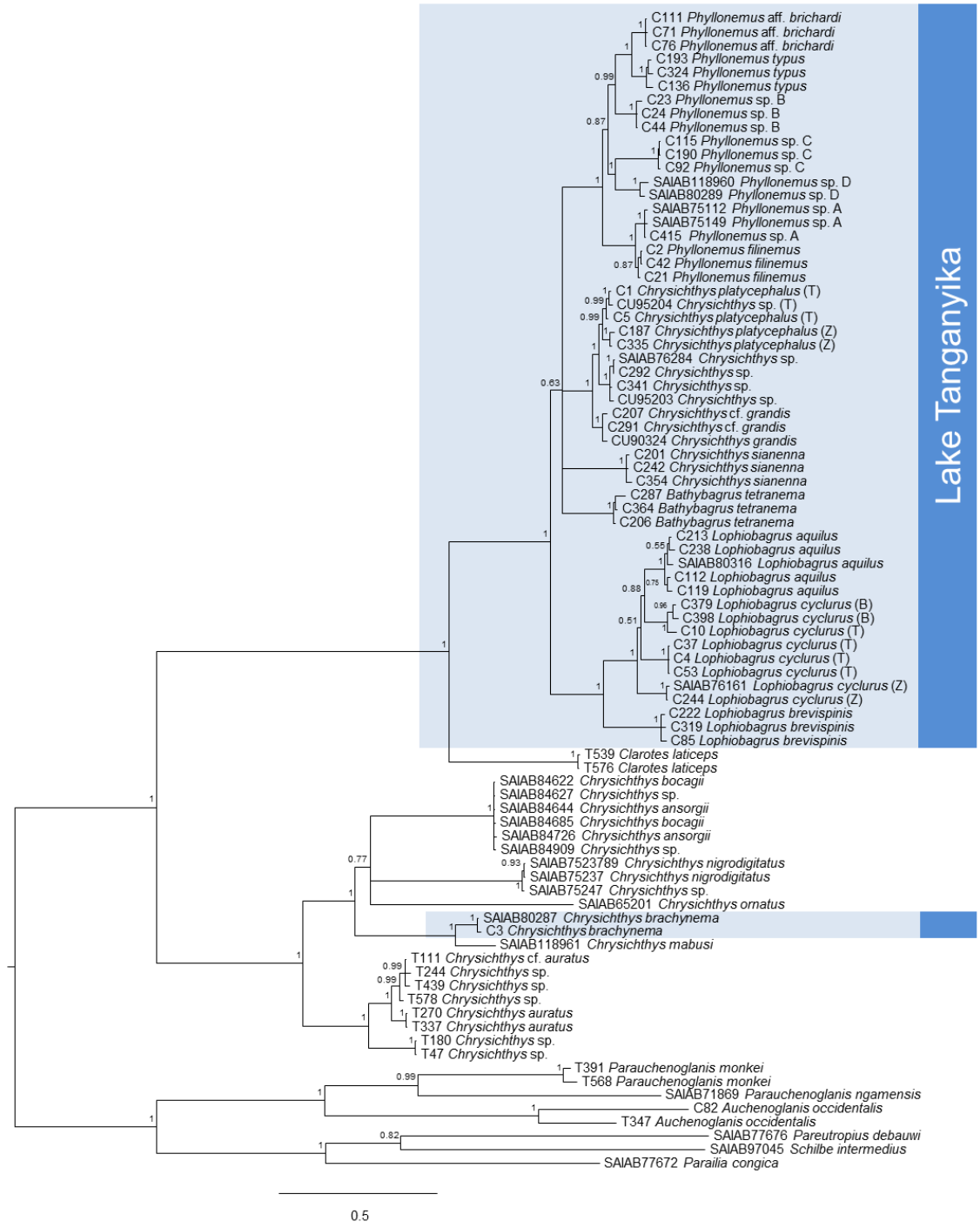


Figure 3 Consensus tree generated from the concatenated mitochondrial MrBayes analysis. Numbers at the nodes are Bayesian posterior probability (BPP) values. Taxa contained within the boxes are those found in Lake Tanganyika.

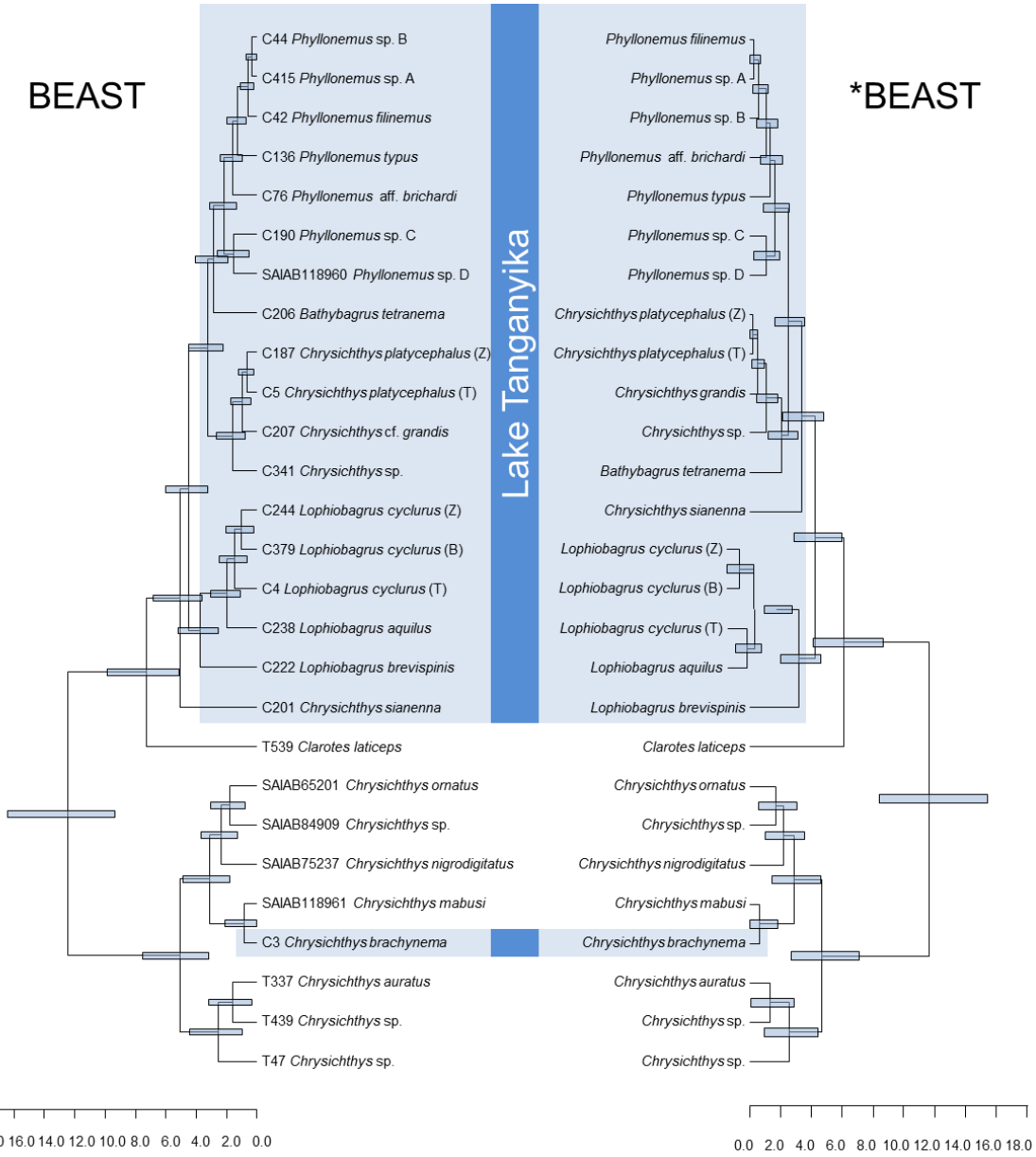
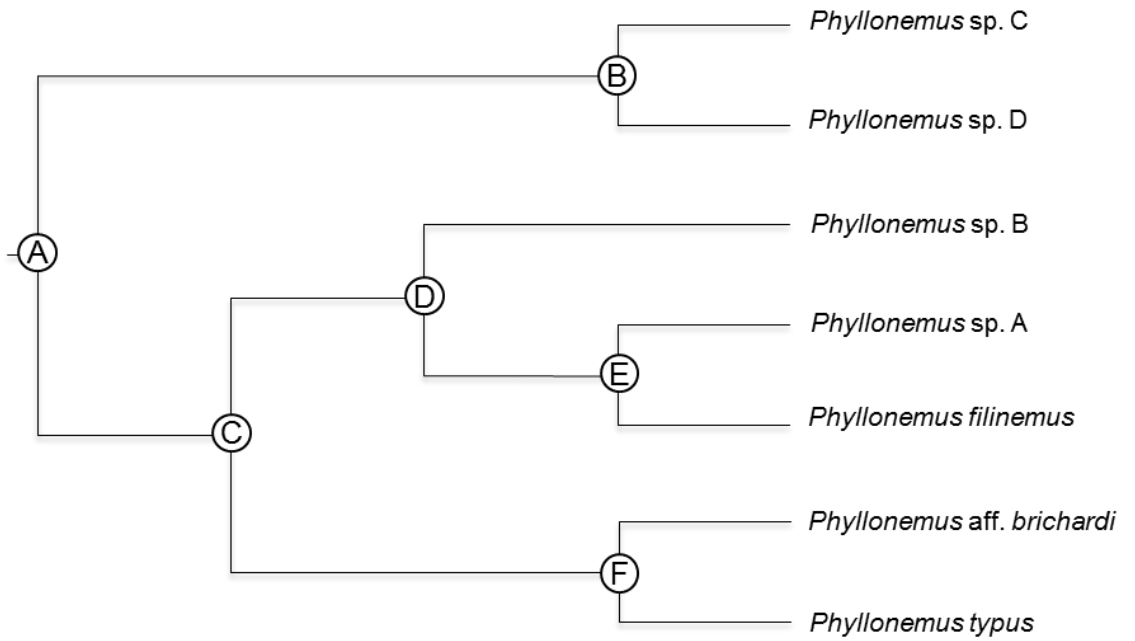


Figure 4 Dated phylogeny generated using BEAST and *BEAST methodologies. Node bars are 95% confidence intervals around node ages. Taxa contained within the two boxes are those found in Lake Tanganyika.

2.4.3 Species delimitation

The different methods of accounting for the different rates of evolution between the protein coding genes and S7 in BP&P (rates drawn from a dirichlet prior and directly specified rates) produced concordant results as did the different species delimitation models (Figure 5). Results of this analysis did not support all tentative morphological assignments and instead five species were supported (using a value of 0.95): *Phyllonemus* sp. A and *Phyllonemus* sp. B are combined with P. filinemus whilst *Phyllonemus* sp. C, *Phyllonemus* sp. D, *P.* aff. *brichardi*, and *P. typus* were supported separately.



Population prior	Locus rate	Model	A	B	C	D	E	F
2, 2000	1	0	1	0.99	1	0.89	0.31	1
2, 2000	1	1	1	0.99	1	0.88	0.31	1
2, 2000	2	0	1	1	1	0.90	0.32	1
2, 2000	2	1	1	1	1	0.90	0.31	1
1, 100	1	0	1	0.98	1	0.80	0.25	1
1, 100	1	1	1	0.98	1	0.82	0.27	1
1, 100	2	0	1	0.98	1	0.81	0.26	1
1, 100	2	1	1	0.98	1	0.81	0.27	1

Figure 5 Guide tree used in the BP&P analysis based on the concatenated MrBayes phylogeny (Figure 1). Posterior probabilities for each node are given in the table. Locus rate 1 specifies the results when the evolutionary rate for each locus was drawn from a dirichlet prior with a = 2. Locus rate 2 specifies rates of evolution for each gene taken from the BEAST analysis. Models 0 and 1 specify the two different species delimitation algorithms available in BP&P

2.5 Discussion

2.5.1 Origins and colonisation history

Bailey and Stewart (1984) hypothesised that the non-*Chrysichthys* LT claroteine genera were each derived from a *Chrysichthys*-like ancestor and noted the clear morphological differences between *C. brachynema* and the other *Chrysichthys* species within LT. However, explicit colonisation scenarios were not proposed because the monophyly of *Chrysichthys* had not been tested using phylogenetic inference. The phylogenetic results do not support monophyly of *Chrysichthys* and instead show that the majority of LT endemic species, with the single exception of *C. brachynema* comprise a clade consistent with a single colonisation leading to the majority of LT claroteine diversity (LT clade), with a separate colonisation event accounting for the presence of *C. brachynema* in LT.

The results suggest that *C. brachynema* has diverged recently at 0.05–2.12 Ma (BEAST HPD) from its inferred sister taxon *Chrysichthys mabusi* from the Chambeshi River (part of the Congo River catchment). However, limited sampling from outside LT means that *C. mabusi* may not be its true sister species suggesting it may be a very young species and that its lack of diversification in LT may be a function of its recent colonisation.

The LT clade diverged from its sister taxon *Clarotes laticeps* much earlier at 5.15–9.88 Ma (BEAST HPD). *Clarotes laticeps* has a widespread distribution from across the Nilo-Sudan region (particularly prevalent in West Africa), but it has not been recorded in any of the countries surrounding LT. This distribution mirrors that seen from the Eocene Mahenge crater lake in Tanzania in that fossils from this site often have their closest relatives in West Africa (e.g., Greenwood and Patterson, 1967; Murray, 2003).

2.5.2 Diversification of the LT claroteine clade

The dated phylogeny suggests that the onset of diversification of the LT claroteine clade occurred between 3.61 and 6.84 Ma (BEAST HPD). This places it well after the formation of Lake Tanganyika that occurred between 9 and 12 Ma, and is suggestive of diversification within full lacustrine conditions that were established around 5–6 Ma (Tiercelin and Mondeguer, 1991). Recent speciation events within the phylogeny suggest that diversification in this clade may be ongoing.

Table 1 Features of Lake Tanganyika radiations taken from the literature.

Taxa	Number of species	Number of colonisations	Onset of diversification (Ma)	Dating methodology	Reference
Synodontis catfishes (Mochokidae)	~11	Single (with one subsequent emigration)	7.9 (5.7-10 HPD)	Relaxed log normal clock in BEAST, single fossil calibration, combined nuclear and mitochondrial data	Day et al., 2013
Mastacembelus eels (Mastacembelidae)	14	Single	7.9 (5.5-10.6 HPD)	Relaxed log normal clock in BEAST single fossil calibration of outgroup, combined nuclear and mitochondrial data	Brown et al., 2010
Platythelphusa crabs (Platythelphusidae)	9	Single	3.3-2.5	Molecular clock taken from marine crabs applied to 16S	Marijnissen et al., 2006
Paludomidae gastropods	~70	Multiple	Predates lake formation by up to 40 Ma	Molecular clock taken from littorinid gastropods applied to 16S	Wilson et al., 2004
Cichlidae (selected tribes)		Dependent on dating method		Novel method involving dates extrapolated from plots of genetic distances against dates estimated from a global phylogeny calibrated either with Gondwanan fragmentation dates or cichlid fossils	Genner et al., 2007
- *Ectodini*	33–41		23.53 ± 4.14 (SD) Gondwanan fragmentation, 10.8 ± 1.94 (SD) fossil		
- *Tropheini*	19–36		6.8 ± 1 (SD) Gondwanan fragmentation, 3.02 ± 0.55 (SD) fossil		
- *Lamprologini*	79–83		16.01 ± 2.44 (SD) Gondwanan fragmentation, 7.33 ± 1.14 (SD) fossil		
- *Limnochromini*	13		10.83 ± 2 (SD) Gondwanan fragmentation, 4.99 ± 0.92 (SD) fossil		
Claroteine catfishes	15 currently described	Two, only one leading to radiation	5.08 (3.61-6.84 HPD) 4.23 (2.88 5.95 HPD)	Single fossil calibration, clock per gene, BEAST and *BEAST methodologies	This study

Divergence estimates within the LT claroteine radiation based on BEAST and *BEAST analyses were broadly congruent, although the latter method generally gave more recent dates. This is consistent with predictions as *BEAST takes into account allelic divergence before speciation (Drummond et al., 2012).

In order to investigate the role of the environment in facilitating diversification it is necessary to compare radiations in disparate clades to investigate temporal patterns across taxa (summarised in Table 1). Many small radiations within LT have resulted from single colonisations (Table 1). The extent to which subsequent emigration has occurred from the LT radiations, as has occurred in *Synodontis*, is currently unclear, and merits investigation through broader sampling from outside LT. The majority of LT radiations have age estimates with confidence intervals that overlap the establishment of full lacustrine conditions around 5–6 Ma (Tiercelin and Mondeguer, 1991); only the *Platythelphusa* crab radiation (Marijnissen et al., 2006) is younger.

One limitation on comparisons between the results of different studies is that different dating methodologies have been employed. In addition, the dating of recent radiations suffers from a lack of suitable calibration points (e.g., Marijnissen et al., 2006) with single calibrations common (Table 1). Singly calibrated phylogenies should be interpreted cautiously due to uncertainty surrounding calibration points as seen in the *Synodontis* radiation where a revised calibration point and much greater taxonomic sampling led to age estimates of the onset of diversification in LT changing from 5.5 Ma (4.0–7.3 HPD) (Day et al., 2009) to 7.9 Ma (5.7–10 HPD) (Day et al., 2013). Working at a broader phylogenetic context allows multiple calibration points to be evaluated and applied. This approach was used by Genner et al. (2007) to date African cichlid radiations but yielded two very different timescales based on fossil calibrations and Gondwanan fragmentation (Table 1).

The only attempt to place the LT claroteines in a broad family level dated phylogeny (based on RAG1 and RAG2) estimated the divergence of *Lophiobagrus* from the *Phyllonemus* and *Bathybagrus* clade at approximately 30 Ma (Lundberg et al., 2007). This date is considerably older than the formation of LT and the estimates found in this study, dated at 3.22–6.01 Ma BEAST (Figure 4). This older divergence time is because the fossil *Chrysichthys mahengeensis* was used to calibrate this part of the siluriform tree. However, its use as a calibration is

problematic given its uncertain identification (see Methods). As this fossil was placed on the node between a single (unidentified) *Chrysichthys* species and *Rheoglanis dendrophorous,* which is currently synonomised with *Chrysichthys* it is possible that this calibration has artificially inflated previously inferred timings of divergences.

2.5.3 Locally restricted diversity

The inclusion of putative new *Phyllonemus* species from several localities comprising sites in Burundi, Tanzania and Zambia suggests that *Phyllonemus* is more diverse than previously thought with five species supported by Bayesian species delimitation. These species appear to be locally restricted, based on the sampling in this study indicating that geographic isolation may be important in driving diversification within this genus. Using a current identification key (Bailey and Stewart, 1984) *Phyllonemus* sp. D was originally identified as *P. filinemus*, however phylogenetic analysis indicates that this taxon does not resolve within or sister to *P. filinemus* but is instead sister to a much smaller new species of *Phyllonemus* with which it occurs in sympatry. This suggests that diversity in *Phyllonemus* is not merely a result of populations diverging neutrally in different basins within LT, but that convergent evolution may play a role, which warrants further investigation. Given the potential for cryptic, locally restricted species within this genus, the lack of genetic samples in this study from Democratic Republic of the Congo is currently highlighted as a major factor in the potential underestimation of species numbers in this genus and it is likely that a detailed survey of this coastline will reveal additional diversity. This study was also limited by relatively low sample sizes in each species and further specimens would increase the power of the analysis to delimit species.

Bayesian species delimitation methods allow multilocus nuclear data to be used to investigate recent speciation events. The reliance on a guide tree limits the applicability of the method for LT claroteines and it is unclear how much of the additional diversity seen in the phylogeny is at the species level or if genetic diversity represents geographical variation. For instance, the different clades within *Lophiobagrus cyclurus* and *Lophiobagrus aquilus* may represent different (mostly geographically isolated) species, however, it is not possible to investigate this hypothesis using BP&P as no guide tree is well supported. The geographic

diversity within *L. cyclurus*, with different clades from Burundi, Tanzania and Zambia, is noteworthy as *L. cyclurus* was originally erected through the synonymisation of *Chrysichthys cyclurus* and *Lophiobagrus lestradei*, from Zambia and Burundi respectively. The existence of locally restricted diversity outlined in this study highlights the importance of lake-wide sampling when investigating LT radiations.

2.5.4 Taxonomy of LT claroteines

The LT clade is well supported in the phylogeny. Within this clade, *Lophiobagrus*, *Phyllonemus* and LT *Chrysichthys* clades are also well supported. Conversely, the placements of *Bathybagrus tetranema* and *Chrysichthys sianenna* are not well supported.

Previously Mo (1991) expanded *Bathybagrus* to include the *Chrysichthys* species within LT without a post-cleithral process (also absent in *Chrysichthys dendrophorus* outside of LT). This taxonomy has not been widely adopted due to concern over the identity of the specimens examined by Mo (Hardman, 2008) and the unique combination of characters present in *B. tetranema*, the only current member of the genus. *Bathybagrus tetranema* exhibits a combination of characters that are present individually in the different genera within LT: presence of subcutaneous eyes (*Lophiobagrus*), absence of post-cleithral process (LT *Chrysichthys*), no nasal barbels (*Phyllonemus*) in addition to no inner mandibular barbels. This study supports the conclusion of Mo that the LT *Chrysichthys* without a post-cleithral process do not resolve within *Chrysichthys* and as such *Chrysichthys* is non-monophyletic. The placement of *B. tetranema* within the LT clade is not strongly supported in any analysis other than the concatenated MrBayes analysis, and this, coupled with its unique combination of features, suggests that it should remain as a separate monotypic genus.

Chrysichthys sianenna also has a unique morphology among the LT claroteines, displaying a highly reduced post-cleithral process (it is not absent as in the other *Chrysichthys* in the LT clade) and numerous (>22) long gill rakers on the first gill arch (Hardman, 2008) suggestive of a planktivorous feeding habit. The phylogenetic placement of *C. sianenna* within the LT clade is uncertain as the mitochondrial and nuclear trees are discordant and its position based on nuclear data is affected by the method of tree reconstruction. *Chrysichthys sianenna* is well

supported within the LT clade, and is therefore not a member of the clade containing the type species of *Chrysichthys*. It does not however resolve within any of the other LT genera or as sister to the other LT *Chrysichthys* in any analysis. In addition, there are significant and distinguishable morphological differences between *C. sianenna* and the remaining LT genera (free orbital rim, the presence of nasal and inner mandibular barbels, forked tail) reinforcing its placement outside of these clades.

Based on the findings of the phylogenetic analyses, the taxonomy of the LT claroteines is in need of revision. A full revision of the LT claroteines, however, requires further morphological analysis to be conducted and is beyond the scope of this study.

2.5.5 Conclusions

There have been two independent colonisations of LT claroteines, one leading to *Chrysichthys brachynema* and the other leading to the remainder of the LT claroteines. This radiation occurred within lacustrine conditions in LT based on fossil calibrated dating of the molecular phylogeny. There is no evidence of subsequent emigration from LT but sampling from the surrounding waterways is sparse. This analysis suggests that the taxonomy of the claroteine needs revision as *Chrysichthys* is found not to be monophyletic. There is also evidence that species richness is underestimated in this radiation with additional species diversity supported in *Phyllonemus* and cryptic diversity also uncovered within *Lophiobagrus*.

2.6 References

Agnèse, J. F., & Teugels, G. G. (2001) The Bathyclarias-Clarias species flock: a new model to understand rapid speciation in African Great Lakes. *Comptes Rendus de l'Académie des Sciences - Series III - Sciences de la Vie,* **324**(8), 683–688

Bailey, R. M., & Stewart, D. J. (1984) Bagrid Catfishes from Lake Tanganyika, with a Key and Descriptions of New Taxa. *Miscellaneous Publications Museum of Zoology, University of Michigan*, **168**

Biomatters (2013) Geneious version 5.6 created by Biomatters. <http://www.geneious.com>

Brandley, M. C., Wang, Y., Guo, X., De Oca, A. N. M., Fería-Ortíz, M., Hikida, T., & Ota, H. (2011) Accommodating heterogenous rates of evolution in molecular divergence dating methods: an example using intercontinental dispersal of Plestiodon (Eumeces) lizards. *Systematic Biology,* **60**(1), 3–15

Brown, K. J., Britz, R., Bills, R., Rüber, L., & Day, J. J. (2011) Pectoral fin loss in the Mastacembelidae: a new species from Lake Tanganyika. *Journal of Zoology,* **284**(4), 286–293

Brown, K. J., Rüber, L., Bills, R., & Day, J. J. (2010) Mastacembelid eels support Lake Tanganyika as an evolutionary hotspot of diversification. *BMC Evolutionary Biology,* **10**, 188

Chow, S., & Hazama, K. (1998) Universal PCR primers for S7 ribosomal protein gene introns in fish. *Molecular Ecology,* **7**(9), 1247–1263

Cohen, A. S., Lezzar, K. E., Tiercelin, J. J., & Soreghan, M. (1997) New palaeogeographic and lake-level reconstructions of Lake Tanganyika: implications for tectonic, climatic and biological evolution in a rift lake. *Basin Research,* **9**(2), 107–132

Cohen, A. S., Soreghan, M. J., & Scholz, C. A. (1993) Estimating the age of formation of lakes: An example from Lake Tanganyika, East African Rift system. *Geology,* **21**, 511

Coulter, G. W. (1991) *Lake Tanganyika and its Life*. Oxford: Oxford University Press.

Danley, P. D., & Kocher, T. D. (2001) Speciation in rapidly diverging systems: lessons from Lake Malawi. *Molecular Ecology,* **10**(5), 1075–1086

Day, J. J., Bills, R., & Friel, J. P. (2009) Lacustrine radiations in African *Synodontis* catfish. *Journal of Evolutionary Biology,* **22**(4), 805–817

Day, J. J., Cotton, J. A., & Barraclough, T. G. (2008) Tempo and mode of diversification of Lake Tanganyika cichlid fishes. *PloS One,* **3**(3), e1730

Day, J. J., Peart, C. R., Brown, K. J., Friel, J. P., Bills, R., & Moritz, T. (2013) Continental diversification of an African catfish radiation (Mochokidae: *Synodontis*). *Systematic Biology,* **62**(3), 351–365

Day, J. J., & Wilkinson, M. (2006) On the origin of the *Synodontis* catfish species flock from Lake Tanganyika. *Biology Letters,* **2**(4), 548–552

Drummond, A. J., Suchard, M. A., Xie, D., & Rambaut, A. (2012) Bayesian phylogenetics with BEAUti and the BEAST 1.7 *Molecular Biology and Evolution,* **29**(8), 1969–1973

Eschmeyer, W. N. (Ed.) (2013) *Catalog of Fishes: Genera, Species, References.* <http://research.calacademy.org/research/ichthyology/catalog/fishcatmain.asp> Accessed on 19th May 2013

Folmer, O., Black, M., Hoeh, W., Lutz, R., & Vrijenhoek, R. (1994) DNA primers for amplification of mitochondrial cytochrome c oxidase subunit 1 from diverse metazoan invertebrates. *Molecular Marine Biology and Biotechnology,* **3**(5), 294–299

Fryer, G., & Iles, T. D. (1972) *The cichlid fishes of the Great Lakes of Africa*. Edinburgh: Oliver and Boyd

Genner, M. J., Seehausen, O., Lunt, D. H., Joyce, D. A., Shaw, P. W., Carvalho, G. R., & Turner, G. F. (2007) Age of cichlids: new dates for ancient lake fish radiations. *Molecular Biology and Evolution,* **24**(5), 1269–1282

Genner, M. J., & Turner, G. F. (2012) Ancient hybridization and phenotypic novelty within Lake Malawi's cichlid fish radiation. *Molecular Biology and Evolution,* **29**(1), 195–206

Greenwood, P. H., & Patterson, C. (1967) A fossil osteoglossoid fish from Tanzania (E. Africa). *Zoological Journal of the Linnean Society,* **47**(311), 211–223

Haesler, M. P., & Seehausen, O. (2005) Inheritance of female mating preference in a sympatric sibling species pair of Lake Victoria cichlids: implications for speciation. *Proceedings of the Royal Society B: Biological Sciences,* **272**(1560), 237–245

Hardman, M. (2005) The phylogenetic relationships among non-diplomystid catfishes as inferred from mitochondrial cytochrome b sequences; the search for the ictalurid sister taxon (Otophysi: Siluriformes). *Molecular Phylogenetics and Evolution,* **37**(3), 700–720

Hardman, M. (2008) A new species of catfish genus Chrysichthys from Lake Tanganyika (Siluriformes: Claroteidae). *Copeia,* **2008**(1), 43

Hardman, M., & Page, L. M. (2003) Phylogenetic relationships among bullhead catfishes of the genus Ameiurus (Siluriformes: Ictaluridae). *Copeia* **2003**(1), 20–33

Harrison, T., Msuya, P., Murray, A., Jacobs, B. F., Ba'ez, A. M., Mundil, R., & Ludwig, K. R. (2001) Paleontological investigations at the Eocene locality of Mahenge in northcentral Tanzania, East Africa. In: Gunnell, G. F. (Ed.), *Eocene Biodiversity: Unusual and Rarely Sampled Habitats.* New York : Kluwer Academic/Plenum Publishers. pp. 39–74

Heled, J., & Drummond, A. J. (2010) Bayesian inference of species trees from multilocus data. *Molecular Biology and Evolution,* **27**(3), 570–580

Huelsenbeck, J. P., & Ronquist, F. (2001) MRBAYES: Bayesian inference of phylogenetic trees. *Bioinformatics,* **17**(8), 754–755

Lanfear, R., Calcott, B., Ho, S. Y. W., & Guindon, S. (2012) Partitionfinder: combined selection of partitioning schemes and substitution models for phylogenetic analyses. *Molecular Biology and Evolution,* **29**(6), 1695–1701

Leaché, A. D., & Fujita, M. K. (2010) Bayesian species delimitation in West African forest geckos (Hemidactylus fasciatus) *Proceedings of the Royal Society B: Biological Sciences,* **277**(1697), 3071–3077

Li, C., Ortí, G., Zhang, G., & Lu, G. (2007) A practical approach to phylogenomics: the phylogeny of ray-finned fish (Actinopterygii) as a case study. *BMC Evolutionary Biology,* **7**, 44

Longbottom, A. (2010) A new species of the catfish Nigerium from the Palaeogene of the Tilemsi Valley, Republic of Mali. *Palaeontology,* **53**(3), 571–594

Lukoschek, V., Scott Keogh, J., & Avise, J. C. (2012) Evaluating fossil calibrations for dating phylogenies in light of rates of molecular evolution: a comparison of three approaches. *Systematic Biology,* **61**(1), 22–43

Lundberg, J. G., Sullivan, J. P., Rodiles-Hernández, R., & Hendrickson, D. A. (2007) Discovery of African roots for the Mesoamerican Chiapas catfish, Lacantunia enigmatica, requires an ancient intercontinental passage. *Proceedings of the Academy of Natural Sciences of Philadelphia,* **156**, 39–53

MacArthur, R. H., & Wilson, E. O. (1967) *The Theory of Island Biogeography*. Princeton: Princeton University Press

Marijnissen, S. A. E., Michel, E., Daniels, S. R., Erpenbeck, D., Menken, S. B. J., & Schram, F. R. (2006) Molecular evidence for recent divergence of Lake Tanganyika endemic crabs (Decapoda: Platythelphusidae). *Molecular Phylogenetics and Evolution,* **40**(2), 628– 634

McCormack, J. E., Heled, J., Delaney, K. S., Peterson, A. T., & Knowles, L. L. (2011) Calibrating divergence times on species trees versus gene trees: implications for speciation history of Aphelocoma jays. *Evolution***, 65**(1), 184–202

Mo, T. P. (1991) *Theses Zoologicae vol. 17: Anatomy, relationships and systematics of the Bagridae (Teleostei: Siluroidei)—with a hypothesis of siluroid phylogeny*. Koenigstein: Koeltz Scientific Books

Murray, A. M. (2003) A new Eocene Citharinoid fish (Ostariophysi: Characiformes) from Tanzania. *Journal of Vertebrate Paleontology,* **23**(3), 501–507

Murray, A. M., & Budney, L. A. (2003) A new species of catfish (Claroteidae, Chrysichthys) from an Eocene crater lake in East Africa. *Canadian Journal of Earth Sciences,* **40**(7), 983–993

Muschick, M., Indermaur, A., & Salzburger, W. (2012) Convergent evolution within an adaptive radiation of cichlid fishes. *Current Biology,* **22**(24), 2362–2368

Nylander, J. A. A., Wilgenbusch, J. C., Warren, D. L., & Swofford, D. L. (2008) AWTY (are we there yet?): a system for graphical exploration of MCMC convergence in Bayesian phylogenetics. *Bioinformatics,* **24**(4), 581–583

Phillips, M. J. (2009) Branch-length estimation bias misleads molecular dating for a vertebrate mitochondrial phylogeny. *Gene,* **441**(1-2), 132–140

Posada, D. (2008) JModelTest: phylogenetic model averaging. *Molecular Biology and Evolution,* **25**(7), 1253–1256

Rambaut, A., & Drummond, A. J. (2009) Tracer v1 5 <http://beast.bio.ed.ac.uk/Tracer>

Risch, L. (1987) Description of four new bagrid catfishes from Africa (Siluriformes: Bagridae) *Cybium* **11**(1), 21–38

Rüber, L., Tassell, J. L. Van., & Zardoya, R. (2003) Rapid speciation and ecological divergence in the American seven-spined gobies (Gobiidae, Gobiosomatini) inferred from a molecular phylogeny. *Evolution,* **57**(7), 1584–1598

Schliewen, U. K., Tautz, D., & Paabo, S. (1994) Sympatric speciation suggested by monophyly of crater lake cichlids. *Nature,* **368**, 629–632

Stewart, K. (1994) A late Miocene fish fauna from Lothagam, Kenya. *Journal of Vertebrate Paleontology,* **14**(4), 592–594

Streelman, J. T., Alfaro, M., Westneat, M. W., Bellwood, D. R., & Karl, S. A. (2002) Evolutionary history of the parrotfishes: biogeography, ecomorphology, and comparative diversity. *Evolution*, **56**(5), 961–971

Sullivan, J. P., Lundberg, J. G., & Hardman, M. (2006) A phylogenetic analysis of the major groups of catfishes (Teleostei: Siluriformes) using rag1 and rag2 nuclear gene sequences. *Molecular Phylogenetics and Evolution,* **41**(3), 636–662

Tiercelin, J. J., & Mondeguer, A. (1991) The geology of the Tanganyika trough. In G. W. Coulter (Ed.), *Lake Tanganyika and its Life.* Oxford: Oxford University Press. pp. 7–48

Wagner, C. E., Harmon, L. J., & Seehausen, O. (2012) Ecological opportunity and sexual selection together predict adaptive radiation. *Nature* **487**(7407), 366–369

Wallace, A. R. (1880) *Island Life or, the Phenomena and Causes of Insular Faunas and Floras: Including a Revision and Attempted Solution of the Problem of Geological Climates*. London: Macmillan and Co.

Ward, R. D., Zemlak, T. S., Innes, B. H., Last, P. R., & Hebert, P. D. N. (2005) DNA barcoding Australia's fish species. *Philosophical Transactions of the Royal Society B: Biological Sciences,* **360**(1462), 1847–1857

West, K., & Michel, E. (2000) The dynamics of endemic diversification: Molecular phylogeny suggests an explosive origin of the thiarid gastropods of Lake Tanganyika. In Rossiter, A., & Kawanabe, H. (Eds.) *Advances in Ecological Research* **31**. pp. 331–354

White, E. (1935) Fossil fishes of Sokoto Province. *Bulletin of the Geological Survey of Niger*, **14**, 1–78

Wilson, A. B., Glaubrecht, M., & Meyer, A. (2004) Ancient lakes as evolutionary reservoirs: evidence from the thalassoid gastropods of Lake Tanganyika. *Proceedings of the Royal Society B: Biological Sciences,* **271**(1538), 529–536

Wright, J. J., & Bailey, R. M. (2012) Systematic revision of the formerly monotypic genus Tanganikallabes (Siluriformes: Clariidae). *Zoological Journal of the Linnean Society,* **165**(1), 121–142

Wright, J. J., & Page, L. M. (2006) Taxonomic revision of Lake Tanganyikan *Synodontis* (Siluriformes: Mochokidae) *Bulletin of the Florida Museum of Natural History*, **46**(4), 99–154

Yang, Z., & Rannala, B. (2010) Bayesian species delimitation using multilocus sequence data. *Proceedings of the National Academy of Sciences*, **107**(20), 9264–9269

Zhang, C., Zhang, D. X., Zhu, T., & Yang, Z. (2011) Evaluation of a bayesian coalescent method of species delimitation. *Systematic Biology,* **60**(6), 747–761

Chapter 3
Intralacustrine allopatry: differing geographic structure in two Lake Tanganyika catfish species

3.1 Abstract

Geographic isolation has been implicated in the divergence of multiple cichlid fish species from Lake Tanganyika through isolation by distance and also through secondary admixture caused by fluctuating lake levels. In non–cichlid taxa, however, even broad scale phylogeographic patterns remain to be elucidated. The catfishes of Lake Tanganyika form multiple replicate radiations and share habitats with cichlids making them a useful comparative system to study the factors influencing and maintaining divergence. Low sample numbers have previously been problematic in studying population divergence, however, restriction site associated (RAD) sequencing provides a large amount of information for each individual from across the genome enabling smaller sample sizes to be used. This study investigates phylogeographic structure throughout Lake Tanganyika in two divergent catfish species with differing biologies: the claroteine, *Lophiobagrus cyclurus*, and the mochokid, *Synodontis multipunctatus*. A strong lake–wide phylogeographic signal is observed in the littoral mouth-brooding *L. cyclurus* including a deep divergence between the northern and southern basins. This divergence is also maintained across a heterogeneous habitat over much smaller distances in the southern basin (78km of Zambian coastline), though analyses suggest there is some on-going admixture between all populations. In contrast the brood parasite *S. multipunctatus,* which was sampled from greater depths, shows little phylogeographic structure with only weak divergence between the northern and southern basins.

3.2 Introduction

The East African Rift Lakes are home to multiple independent endemic radiations in a wide variety of taxa, and as such offer the opportunity to compare the relative roles of different factors that influence diversification. Multiple factors have been highlighted as important in these environments, including competition (Winkelmann et al., 2014), niche partitioning (e.g., Genner et al., 1999; Marijnissen

et al., 2008), sensory drive (Seehausen et al., 2008), hybridisation (Genner and Turner, 2012; Salzburger et al., 2002) and intralacustrine allopatry (e.g., Koblmüller et al., 2011; Verheyen et al., 1996) but the extent to which these factors apply to the different radiations remains to be studied. Extrinsic environmental factors (e.g., habitat barriers, lake level fluctuations) have the potential to interact with species-specific characteristics (e.g., breeding strategy, ecological specialisation) to affect diversification potential in each species and similarities between patterns of genetic structure across multiple taxa indicate the importance of their shared history. Comparisons of phylogeographic structure and the role of habitat barriers in species from multiple radiations will allow generalisations to be investigated between disparate taxa with different phylogenetic constraints, ecologies and behaviours allowing a greater understanding of the processes driving diversification in these environments.

Lake Tanganyika (LT) is the oldest (9-12 Ma, Cohen et al., 1993), deepest (1470m) and longest of the East African Rift lakes (Figure 1). The lake is formed of multiple basins and those in the north and south of LT formed later than the central basin (7-8 Ma and 2-4 Ma respectively, Cohen et al., 1993) leading to a complex history of varying connectivity and lake size. During the more recent late Pleistocene glaciations lake levels dropped ≈ 435m prior to ≈ 106 Ka and while the size of LT was considerably reduced, it remained a large and mostly connected water body (McGlue et al., 2008). The signatures of these Pleistocene lakeshores are suggested from modern distributions of mtDNA lineages in eretmodine cichlids (Rüber et al., 1999; Verheyen et al., 1996). Broad scale patterns do differ between cichlid taxa, for example in the rocky shore genus *Tropheus* some mitochondrial lineages are geographically restricted whereas others are widespread (Baric et al., 2003) and mitochondrial signatures are not always congruent with AFLP data (Egger et al., 2007), whereas in the closely related genus *Simochromis* there is little structure over large distances (300km, Meyer et al., 1996). Phylogeographic structure is also absent in the benthopelagic giant cichlid, *Boulengerochromis microlepis* (Koblmüller et al., 2014).

In some rocky shore cichlid species short breaks in this habitat can act as a barrier to dispersal, for example, across the 7km of Mbete Bay in Zambia genetic structure is seen in a variety of cichlid species e.g., *Tropheus moorii*, *Eretmodus cyanostictus* and *Ophthalmotilapia ventralis* (Sefc et al., 2007). Patterns of

population isolation are influenced by lake level fluctuations which have been implicated in the diversification of LT cichlid taxa, primarily through repeated periods of isolation followed by secondary contact (Egger et al., 2007; Nevado et al., 2013). The diversity of geographic patterns in cichlids suggest a complex interplay of factors influencing allopatric divergence, yet currently even the broad scale geographic patterns of taxa from non-cichlid radiations are poorly understood. It remains to be investigated whether patterns similar to those found in cichlids apply to other less diverse taxa without the high rates of speciation and morphological change seen in cichlids (Rabosky et al., 2013).

The catfishes of LT represent a useful comparative study system in which to study these patterns, as they encompass multiple independent radiations and a wide range of ecologies, from those found in deep water to benthic species from the littoral zone, which would be expected to be most vulnerable to displacement by lake level changes. There are three endemic catfish radiations in LT, in the genera *Tanganikallabes* (three species), and *Synodontis* (≈ 11 species), and in the subfamily Claroteinae (15 species described with additional diversity, Chapter 2; Peart et al., 2014). The distributions of many non-cichlid species are poorly understood with many species assumed to have lake-wide distributions, however, molecular phylogenetics has revealed cryptic diversity in *Synodontis* (Day and Wilkinson, 2006; Koblmüller et al., 2006) and several claroteine genera (Peart et al., 2014). In addition, *Tanganikallabes* and *Synodontis* species have also been described from single localities (Wright and Bailey, 2012; Wright and Page, 2006) suggesting that allopatry may play at least some role in diversification. Studies of intraspecific geographic patterns in these taxa have not yet been conducted and will allow investigation into the possibility of nascent diversification via geographic restriction and comparisons with other widespread taxa in LT. This study focuses on two catfish species with lake-wide distributions: *Lophiobagrus cyclurus*, in which genetic structure between geographically distant populations is observed in a combined nuclear and mitochondrial dataset (Chapter 2; Peart et al., 2014) and *Synodontis multipunctatus* which in contrast shows no geographic structure in a phylogeny based on mitochondrial *Cytb* data (Appendix 2, Figure 1). Both of these taxa have been assessed by the IUCN red list as least concern (Ntakimazi, 2006a, 2006b) based on the fact that although there are local threats (siltation, fisheries) there are no widespread threats throughout their range,

however, the level of population restriction and what constitutes a habitat barrier in these taxa is unknown.

The breeding behaviours and ecologies of the two study species also add a further comparative dimension to this study. *Lophiobagrus cyclurus* is a mouth brooder (Ochi et al., 2002), a trait which might be expected to reduce dispersal ability through extended parental care and is also found in a number of cichlid taxa for which phylogeographic studies have been undertaken (e.g., Rüber et al., 1999; Sefc et al., 2007; Wagner and McCune, 2009). *Synodontis multipunctatus* is a brood parasite of multiple mouth brooding cichlid species (Sato, 1986) including species that have been shown to have differential structure across short geographic distances (Koblmüller et al., 2011; Sefc et al., 2007). In terms of depth *Lophiobagrus cyclurus* is thought to be restricted to the littoral zone (Bailey and Stewart, 1984), potentially increasing its susceptibility to lake level changes in terms of its geographic distribution. *Synodontis multipunctatus* is also found in the littoral zone but has been sampled at depths of up to 170m (Coulter, 1991). In addition to samples of *L. cyclurus* its sister species *L. aquilus* is also included here in order to investigate introgression in the *Lophiobagrus* dataset. This species is found in the same rocky shore environment as *L. cyclurus* and is a paternal mouth brooder (Ochi et al., 2002).

Population level studies using traditional genetic markers require a large amount of data in terms of large numbers of individuals as well as many sampling sites in order to provide sufficient resolution. In species where sampling is limited, or specimens are difficult to collect in large numbers, as is the case for some of the poorly studied non-cichlid LT taxa, this has been a barrier to intraspecific studies. Restriction site associated (RAD) sequencing provides genome-wide genetic data, with a large number of loci. The far higher number of loci per individual means that similarly robust estimates of divergence can be calculated using far fewer individuals (e.g., Willing et al., 2012), overcoming the restriction imposed by reduced sampling. In addition, genome wide genetic data can provide evidence for past and current hybridisation (e.g., Hohenlohe et al., 2011; Keller et al., 2013; Nadeau et al., 2013) and avoids problems associated with markers that are inherited uni-parentally. Here sequence data from RAD loci are used to compare population structure in the two catfish species, *L. cyclurus* and *S. multipunctatus* to address the following questions i) Is there any evidence of lake-wide

phylogeographic structure? ii) How does this pattern vary in direction or strength between the focal species? iii) Is there localised population structure in *L. cyclurus* populations that reveal isolation by distance in areas with potential habitat barriers? iv) Is there evidence for introgression in *Lophiobagrus*?

3.3 Methods

3.3.1 Sampling

A total of 64 specimens were included in this study and collected using multiple methods dependent on locality, including scuba, snorkelling and rotenone (Robertson and Smith-Vaniz, 2008). In addition, *S. multipunctatus* specimens were also collected using gill nets as part of the Department of Fisheries, Mpulungu sampling programme. For the *Lophiobagrus* dataset 33 specimens of *L. cyclurus* including eight specimens each from Burundi, Kigoma (Tanzania), Mpulungu (Zambia) and nine from Sumbu (Zambia) with one individual subsequently filtered out, and an additional seven specimens of *L. aquilus* from both Sumbu and Mpulungu were included. A total of 24 *S. multipunctatus* specimens were included with eight specimens each from Burundi, Kigoma and Mpulungu. Sampling localities are shown in Figure 1. Samples were stored in ethanol prior to extraction.

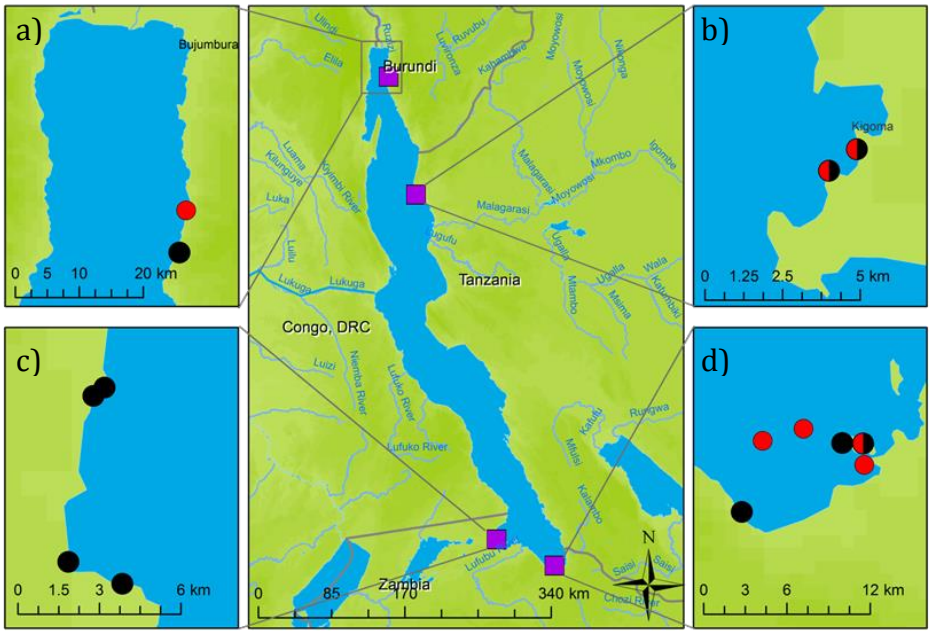


Figure 1 Map (made in ESRI ArcMap10) showing sampling locations for *Synodontis multipunctatus* (red) and *Lophiobagrus cyclurus and aquilus* (black). Sites where species from both genera were taken are shown as half red, half black. Samples are taken from four locales around Lake Tanganyika, a) Burundi, b) Kigoma, c) Sumbu, d) Mpulungu.

3.3.2 Molecular methods

The protocol for constructing RAD libraries was modified from Baird et al (2008) and is described below. DNA was extracted from fin clips using a Qiagen DNeasy kit following the tissue protocol with the addition of an RNase A step to remove RNA. DNA was run on a 0.8% agarose gel to assess the size of DNA fragments. The 260/280 ratio was assessed using a ND 8000 NanoDrop. DNA was quantified using a Qubit 2.0 Fluorometer (Invitrogen). DNA (800ng per individual) was digested with the enzyme Sbf1-HF (20 units, New England BioLabs) which uses the cut site 5' CCTGCA|GG 3' at 37°C. This reaction was purified using Agencourt AMPure XP beads following the Illumina Truseq AMPure XP bead protocol. P1 adaptors, which contain the forward amplification primer site, an Illumina sequencing primer site and an individual barcode, were ligated to each sample of digested DNA. The adaptors have the following sequences where 'x' shows the barcode site and TGCA ligates to Sbf1 digested DNA.

Top:

AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCTxxxxxTGCA

Bottom:

xxxxxAGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGTAGATCTCGGTGGTCGCCGTATCATT.

Following this step, samples were pooled into four libraries each containing 16 individuals (barcodes and library configuration are given in Appendix 2, Table 1). DNA samples were sheared to produce fragments less than 1kb in length using a Covaris S2 sonicator for 40 seconds followed by a gel extraction using a MinElute Gel Extraction kit (Qiagen) to size select fragments between 300-700bp.

A Quick Blunting Kit (New England BioLabs) was used to perform end repair by producing blunt phosphorylated ends from overhangs created by shearing. This reaction was then purified using the AMPure XP bead protocol as above. Following this an 'A' base was added to the 3' blunt ends in order to prepare the fragments for the ligation of the P2 adaptor using Klenow Fragment (New England BioLabs). This reaction was then purified using the AMPure XP bead protocol as above. The P2 adaptors were then ligated, these adaptors have divergent ends containing the reverse complement of the reverse amplification primer site, with the following sequences:

Top:

CTCAGGCATCACTCGATTCCTCCGAGAACAA

Bottom:

CAAGCAGAAGACGGCATACGACGGAGGAATCGAGTGATGCCTGAGT

This prevents amplification of fragments without a P1 adaptor in the following PCR step. This reaction was then purified using AMPure XP beads as above. PCR reactions were completed using Phusion High-Fidelity Master Mix (New England BioLabs) with an initial temperature of 98°C for 30 seconds followed by multiple cycles of 10 seconds 98° C, 30 seconds 65° C, 30 sec 72° C, and a final five minute extension phase at 72° C. Initial trials were completed with 12, 14, 16, 18 and 20 cycles before 17 cycles using 2ul of temple per 10µl reaction were selected as the conditions for larger volume amplification. This resulted in 200µl of PCR product that was concentrated using AMPure XP beads into 20µl before a final gel extraction step to remove contaminant bands was performed using a MinElute Gel Extraction kit (Qiagen). Each library was then quantified using a Qubit 2.0 Fluorometer (Invitrogen) before being sent to FAS Center for Systems Biology, Harvard University for further quantification by qPCR and sequenced across two lanes of 100bp paired end Illumina Hi-seq (two libraries per lane).

3.3.3 Data processing-SNP calling and filtering

Preliminary processing of the data was conducted using programs from the Stacks pipeline v. 1.08 (Catchen et al., 2011). Sequences from each Hi-seq lane were demultiplexed using the process radtags command. This command removes reads with any uncalled bases, discards low quality reads and removes reads where the cut site and barcodes are not present (allowing for some error). The output of this command was combined with sequences from a previous run for five *Lophiobagrus* individuals C228, C134, C245, C171 and C236. Following this the *S. multipunctatus* and *Lophiobagrus* datasets were analysed separately, with the following instructions repeated for each dataset. Fastq files for the first reads for each individual were concatenated to produce a superparent. Following this denovo_map.pl was run on the first read files using a superparent approach. This approach puts all sequences from all individuals into a single file and uses this file to create a catalog that each sample is matched against. The program

sort_read_pairs,pl was used to collate the paired-end reads for each catalog locus (identified in the previous step) that were present in at least two individuals and the superparent. These files were assembled with Velvet (Zerbino and Birney, 2008) by using exec_velvet.pl that creates a fasta file in which each line corresponds to an assembly of the second reads. All tags with more than one assembled contig were eliminated to avoid including instances where several tags had been collapsed into one tag in the clustering process, which ignores the second read. These single contigs were reverse complemented and along with the tags from the first read used to generate a reference with the first read connected to the second read contigs by a string of Ns using a perl script (Hoffman et al., 2014).

The Burrows-Wheeler algorithm was used to map the reads from each individual (first and second) to the reference sequence using the BWA-MEM algorithm (http://bio-bwa.sourceforge.net/). Duplicates were removed using the MarkDuplicates.jar in picard-tools-1.102 (available http://picard.sourceforge.net- no paper). Local realignment and SNP calling was performed with GenomeAnalysisTK (GATK) (DePristo et al., 2011; McKenna et al., 2010). Local realignment using the GATK IndelRealigner is necessary to correct misalignments caused by insertions or deletions. The GATK UnifiedGenotyper was used to call SNPs across all samples using default parameters with the exception of heterozygosity, which was assigned the value of 0.01. The resultant vcf file was filtered to include only high quality calls with the following parameters, SNP quality score of 30, genotype quality score of 20, mapping quality score of 20, sites with coverage under five were excluded. In addition, in an attempt to avoid including repeated transposon regions in all datasets, sites with over 95% of coverage were excluded, as were those with multiple alleles at the same location (perl script, Hoffman et al., 2014). Sites that were present in at least two individuals were included as it has been shown that stringent exclusion of missing data can bias the dataset (Huateng and Knowles, 2014). This resulted in very few called sites for the individual C137 from the *Lophiobagrus* dataset and so this individual was excluded from further analysis.

3.3.4 **Population structure**

Population structure was investigated using Bayesian clustering in STRUCTURE 2.3 (Pritchard et al., 2000) using a matrix of variable sites. STRUCTURE assumes no

linkage disequilibrium so to account for this only one SNP per contig was retained. Structure in *Lophiobagrus* was investigated both with and without *L. aquilus* leading to datasets of 23,677 sites for the combined *L. cyclurus* and *L. aquilus* dataset, 22,204 sites for the *L. cyclurus* only dataset and 17,009 for the *S. multipunctatus* dataset. Values of K from one to eight were investigated using a model of admixture with correlated allele frequencies (Falush et al., 2003), with five replicates per K value with 100,000 generations as burn-in before a further 100,000 generations were sampled. Three replicates (K 1-8) were also completed using the admixture model without correlated allele frequencies but this was found to resolve less structure and the model with correlated allele frequencies was retained. The runs were collated using STRUCTURE HARVESTER (Earl and vonHoldt, 2011) to calculate the mean value of ln Pr(X|K) for each K value and to calculate ΔK (Evanno et al., 2005) which is based on the rate of change between the ln Pr(X|K) for successive values of K. Independent runs for the suggested K values were combined and averaged in CLUMPP (Jakobsson and Rosenberg, 2007) and the output visualised in distruct (Rosenberg, 2004).

Structure within the *L. cyclurus* and *S. multipunctatus* datasets was further investigated using principal component analysis (PCA) in the R packages 'adegenet' (Jombart et al., 2008) and ade4 (Dray and Dufour, 2007). Missing data can influence PCA results if the missing data shows population structure so PCA analyses were conducted on datasets of variable sites present in all individuals using only one SNP per contig leading to datasets of 2,065 loci for *L. cyclurus* and 5,116 loci for *S. multipunctatus*.

In order to investigate differentiation between populations, F_{ST} values were calculated. Population sizes were small so the index of Reich et al. (2009) was employed as this performed well on small sample sizes in a recent simulation study (Willing et al., 2012). Analyses were completed using one SNP per contig with code modified from Rheindt et al. (2014). Confidence intervals for these Fst values were calculated using the jackknife method, following the calculation of Fst using the whole dataset, a set of pseudo Fst values were calculated by omitting one of the individuals across the two populations in turn. The variance and 95% confidence intervals were calculated from the difference between the whole sample estimate of Fst and the pseudo-value estimate. The Fst values calculated

using this method were compared to those generated using the formula of Hudson et al. (1992) in the python library egglib (De Mita and Siol, 2012).

3.3.5 Phylogenetic analysis

Maximum likelihood (ML) trees were calculated with the GTR +GAMMA model using RAxML 7.7.8 (Stamatakis, 2006) with 1,000 non-parametric rapid bootstraps. These analyses were performed on alignments containing both variant and non-variant sites: 2,628,515 sites for the *Lophiobagrus* dataset and 3,356,174 sites for the *S. multipunctatus* dataset.

3.3.6 Heterozygosity estimates

Heterozygosity estimates were calculated in order to investigate if isolation led to a loss of heterozygosity in any of the populations. Hoffman et al. (2014) found that RAD estimates of heterozygosity were less correlated with pedigree based estimates with a combination of increased genotype quality filtering and decreasing low coverage threshold. To account for this, heterozygosity was calculated using a dataset with less stringent filtering, a genotype quality score and low coverage filter of ten. Individual heterozygosity was calculated as the number of heterozygous sites divided by the total number of sites for which an individual was called (Hoffman et al., 2014). Differences between populations were assessed through ANOVA calculations.

3.3.7 Introgression tests

In the *Lophiobagrus* dataset admixture between populations of *L. cyclurus* was investigated using the ABBA/BABA test (Green et al., 2010; Durand et al., 2011). This test is based on four taxon trees with the structure (((1, 2), 3), 4) and the test investigates the relative proportion of ABBAs (a pattern where populations 2 and 3 share the variant allele whereas populations 1 and 4 have the ancestral allele) and BABAs (a pattern where populations 1 and 3 share the variant allele and populations 2 and 4 share the ancestral allele). In this case *L. aquilus* was always used as the outgroup (population 4). ABBA and BABA patterns can arise from either gene flow or incomplete lineage sorting, but under incomplete lineage sorting the proportions of ABBA and BABA patterns would be predicted to be equal. The D-statistic is a measure of how the ratio of ABBA/BABA patterns differs

from equality with a positive D-statistic representing an excess of ABBA over BABA sites. The significance of the D-statistic (from 0) is usually calculated by computing the standard error of the D-statistic using a block jack-knife approach over linkage groups. This approach was not possible in this case as in the absence of a linkage map or reference genome, the order of RAD contigs is unknown. In an attempt to get some measure of the reliability of the D-statistic the dataset was subsampled randomly 1,000 times including 99%, 95%, 90%, 80% and 70% of the data and calculating the mean and standard error of the D-statistic under these conditions to investigate the strength of the difference of the statistic from zero. This analysis was repeated using the tree supported by the RAxML analysis with the following four taxon combinations, Tree 1: (((Burundi, Kigoma), Mpulungu), *L. aquilus*), Tree 2: (((Burundi, Kigoma), Sumbu), *L. aquilus*), Tree 3: (((Mpulungu, Sumbu), Burundi), *L. aquilus*) and Tree 4: (((Mpulungu, Sumbu), Kigoma), *L. aquilus*). These analyses were conducted using a modified version of a python script from Rheindt et al. (2014). The potential introgression between the *L. aquilus* sample C228 and *L. cyclurus* (Figure 2) is a possible confounding factor so these analyses were repeated without this sample.

3.3.8 Relationship with geographic distance

The statistical significance of the relationship between geographic distance and genetic distance was calculated using Mantel tests performed with 999 permutations in the R package ade4 (Chessel et al., 2004). Straight line distances between geographic sites were calculated using the haversine method for calculating distances on the surface of a sphere. One *S. multipunctatus* specimen was excluded from these analyses as it was collected from a gill nets survey by the Department of Fisheries in Zambia at 60m depth for which GPS data was not available. In addition, distances along the lakeshore were calculated using ERSI ArcMap 10's Network Analyst package using the Detailed Water Bodies layer. Lakeshore distances were only calculated for *L. cyclurus* as this taxon was sampled only from the littoral zone (<15m) so might be expected to be more influenced by lakeshore distance than straight line distances whereas *S. multipunctatus* was also collected at up to 100m depth in gill nets. The analysis was performed using average genetic distance between collection localities as samples from the same

collection locality are pseudoreplicates. The uncorrected P-distances were calculated using the R package ape (Paradis et al., 2004).

3.4 Results

3.4.1 RAD-seq characteristics

After the data had been demultiplexed, 303 million reads were retained that passed the quality filter, and contained barcodes and restriction sites, leading to an average of 4.7 million reads per individual. The RAD reference genome based on contigs present in at least two individuals for the *Lophiobagrus* dataset consisted of 33,833 contigs and 16,824,792 base pairs. Following filtering the final *Lophiobagrus* alignment consisted of 2,628,515 sites of which 299,633 were variable between at least two individuals and 266,344 of these were variable within *L. cyclurus*. The RAD reference genome for the *S. multipunctatus* dataset consisted of 26,472 contigs and 13,323,869 base pairs. After filtering, the final *S. multipunctatus* alignment consisted of 3,356,174 sites of which 107,321 were variable.

3.4.2 Variable divergence patterns in *L. cyclurus* and *S. multipunctatus*

In the STRUCTURE analysis the lowest mean value of ln Pr(X|K) for the *Lophiobagrus* dataset including *L. aquilus* was K=5 (Figure 2). At this value of K the samples of *L. aquilus* form a separate cluster with the exception of a single individual, C228, which is derived almost equally from both the *L. aquilus* and Sumbu clusters. Within the Zambian populations, the Sumbu samples resolve mainly in one cluster (coloured red in Figure 2) with the exception of a single individual, C171. This sample resolves fully in a cluster (coloured blue in Figure 2) that is also found in samples from Mpulungu, which are all mixed with the Sumbu cluster. The samples from Burundi and Kigoma are assigned to separate clusters, with a very small amount of the Kigoma cluster in the Burundi samples. There is also some variation in sample C12 (Kigoma) from the other *L. cyclurus* clusters but this sample has a relatively high level of missing data. The Evanno method selected K=3 which separates out the northern and southern basins for *L. cyclurus* and shows the same pattern for *L. aquilus* with C228 split with the cluster from the southern basin. (Appendix 2, Figure 2).

In the *L. cyclurus* only analysis the lowest mean value of ln Pr(X|K) was K=4 which showed the same pattern for *L. cyclurus* as described above (Appendix 2, Figure 3). The Evanno method selected K=3 which separated out the populations from Burundi and Kigoma but assigned all the samples from Zambia to a single cluster (Appendix 2, Figure 4), the same pattern as seen in the K=4 analysis including *L. aquilus*.

In the *S. multipunctatus* STRUCTURE analysis the lowest mean value of ln Pr(X|K) was for K=1 supporting no structure while K=3 was selected by the Evanno method. This value of K separates the northern and southern basins with an additional cluster spread throughout the samples as a minor constituent (Appendix 2, Figure 5), K=2 separates the Zambian population from those in the northern basin without the additional noise (Figure 2).

In the PCA of *L. cyclurus* the first two PCA axes correspond to 28.36% and 20.2% of the variation respectively. The PCA plot (Figure 3) shows three distinct clusters along the first two PCA axes corresponding to the samples from Kigoma, Burundi and Zambia (samples from Mpulungu and Sumbu combined). In the *S. multipunctatus* dataset the first two PCA axes represent 29.53% and 19.03% of the variation respectively. The PCA plot (Figure 3) has three clusters corresponding to the three different populations though these are less tightly clustered than the *L. cyclurus* dataset.

The Fst values for *Lophiobagrus* and *S. multipunctatus* calculated using the formula of Reich et al. (2009) and the egglib python library were broadly concordant, though the values calculated using the egglib library were consistently slightly larger (Appendix 2, Table 2). Within *Lophiobagrus*, the Fst values between the Zambian populations (Mpulungu and Sumbu) and Kigoma (Tanzania) are larger than those between the Zambian populations and Bujumbura (Burundi), which is further away geographically. The Fst value between the two Zambian populations is 0.0341 (formula from Reich et al., 2009) showing some divergence between these populations (Table 1). In the *S. multipunctatus* dataset, the Fst values calculated using both methods were below zero, with the single exception of Mpulungu-Kigoma using the egglib library at 0.0009 (Appendix 2, Table 3) indicates a lack of population differentiation in this species (Table 1).

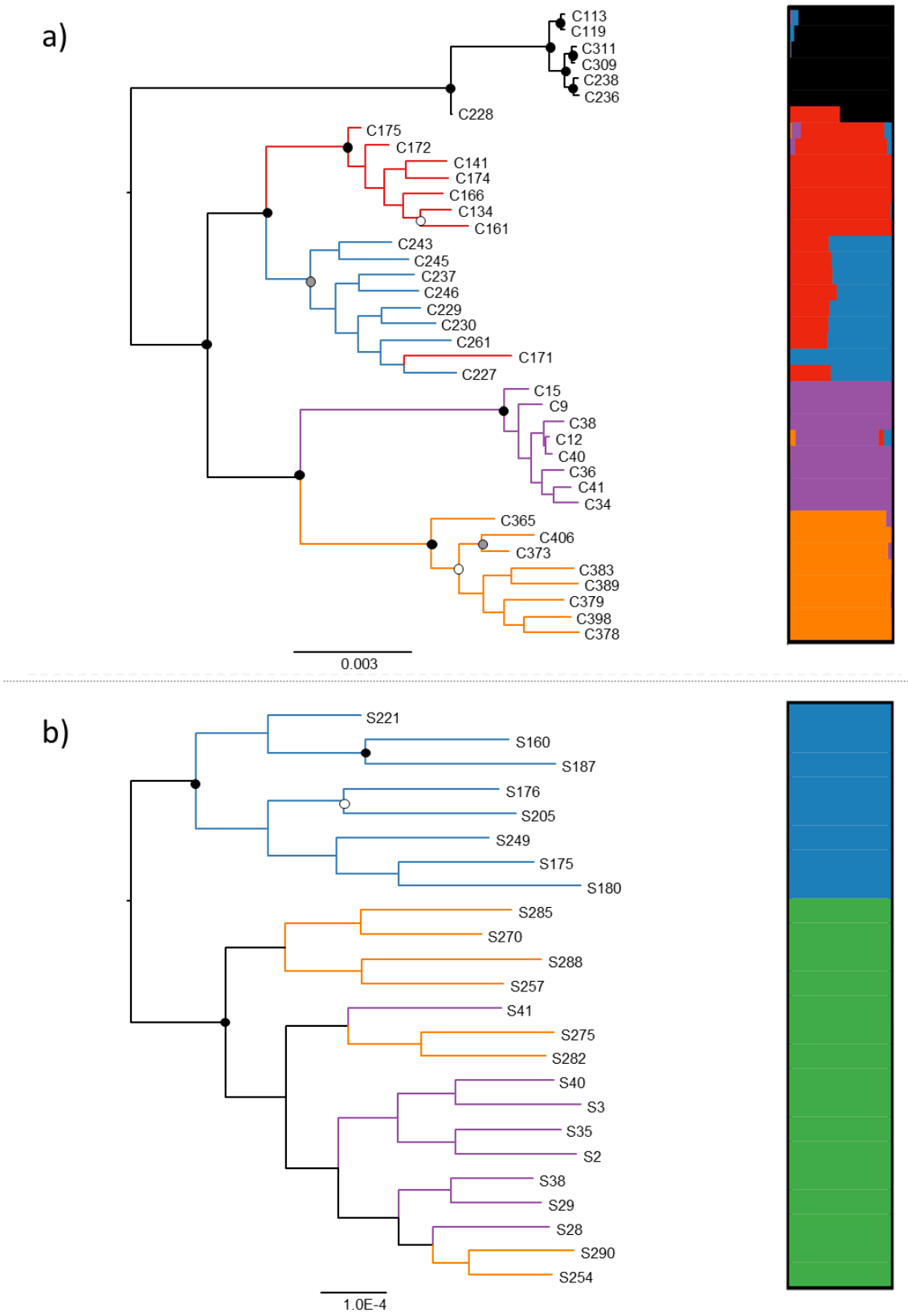


Figure 2. RAxML trees (bootstrap support black circles 100%, grey circles >90%, white circles >80%) and STRUCTURE plots for a) *Lophiobagrus*, b) *S. multipunctatus*. Colours in the trees show collection locality of the specimen: Blue-Mpulungu, Red-Sumbu, Purple-Kigoma, Orange- Burundi. Black branches in the *Lophiobagrus* tree (a) are *L. aquilus*. Green in the STRUCTURE plot (b) combines both sites from the northern basin.

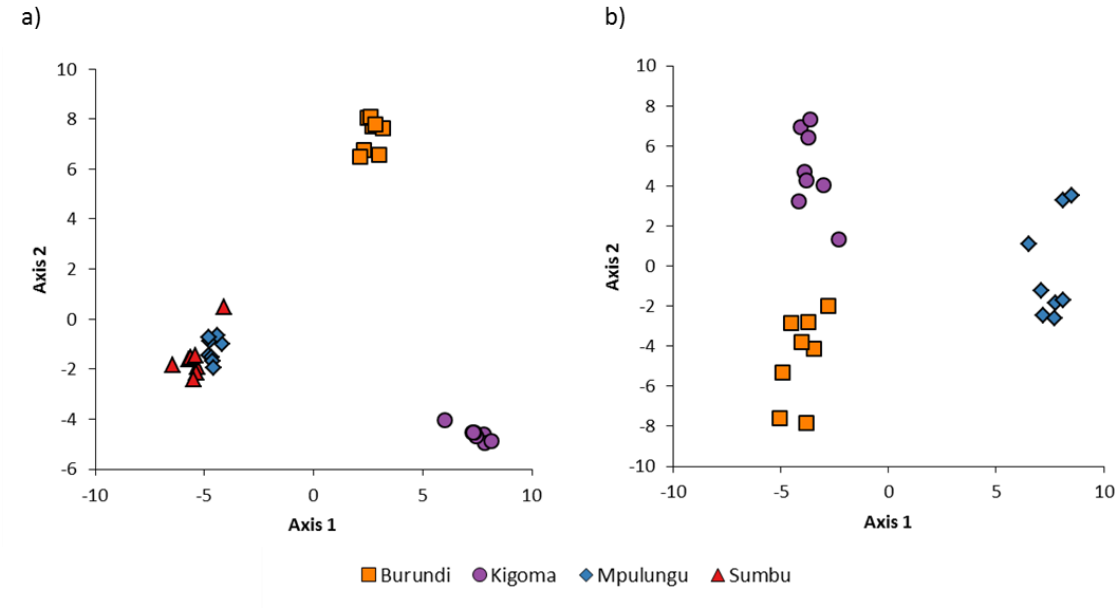


Figure 3 PCA plots for (a) *L. cyclurus* and (b) *S. multipunctatus* populations. Populations in both genera are represented by the following colours: Mpulungu - Blue; Sumbu – Red; Bujumbura - Orange; Kigoma - Purple. There are no *S. multipunctatus* samples from Sumbu.

Table 1 Fst values for (a) *Lophiobagrus* and (b) *S. multipunctatus* (Index from Reich et al. 2009). Numbers in brackets are 95% confidence intervals as calculated by the jackknife method.

(a)

	Burundi	**Kigoma**	**Sumbu**	**Mpulungu**
Kigoma	0.178 (0.170 – 0.181)			
Sumbu	0.170 (0.159 – 0.175)	0.198 (0.188 – 0.205)		
Mpulungu	0.162 (0.154 – 0.165)	0.191 (0.183 – 0.195)	0.034 (0.024 – 0.040)	
L. aquilus	0.307 (0.264 – 0.350)	0.264 (0.199 – 0.333)	0.264 (0.199 – 0.333)	0.255 (0.200 – 0.313)

(b)

	Burundi	**Kigoma**
Kigoma	-0.073 (-0.084 – -0.069)	
Mpulungu	-0.039 (-0.051 – -0.035)	-0.040 (-0.051 – -0.036)

In the *Lophiobagrus* dataset individual heterozygosity is highest for the *L. aquilus* sample C228, which shows mixed assignment in the STRUCTURE analysis but in general was lower for *L. aquilus* than for the *L. cyclurus* populations (Figure 4). When C228 was removed the difference between *L. aquilus* and the *L. cyclurus* populations is significant at the 95% level (ANOVA, $F=22.651$, $p=<0.001$) with Tukey's HSD test indicating that this is mainly influenced by significant differences between *L. aquilus* and each *L. cyclurus* population. When only *L. cyclurus* samples are included the significant relationship remains (ANOVA, $F=8.058$, $p<0.001$) with Tukey's HSD test indicating that the differences between Kigoma and Burundi ($p<0.001$), and Kigoma and Mpulungu ($p=0.002$) are significant. In the *S. multipunctatus* dataset heterozygosity did not vary significantly between populations (ANOVA, $F=1.677$, $p=0.211$; Figure 4).

The ML tree for the *Lophiobagrus* dataset supports *L. cyclurus* as monophyletic (100%, Figure 2), which was not seen in previous analyses (Chapter 2; Peart et al. 2014). In addition, much longer branches separate *L. aquilus* and *L. cyclurus* compared to internal branch lengths between *L. cyclurus* populations. There is a strong geographic pattern with strong support (100%) for the separation of taxa from the northern and southern basins, while within the northern basin samples from Burundi and Kigoma comprise separate clades (100%). The Zambian samples also resolve in different clades based on collection site (100%) with the exception of C171, a specimen from Sumbu that is found in the clade containing Mpulungu samples (Figure 2), which reflects the results of the STRUCTURE analyses (Figure 2). In the *S. multipunctatus* ML tree there is strong support (100%) for the separation of taxa from the northern and southern basins, but there is no support for further structure within the northern basin (Figure 2). The branch lengths between the northern and southern basins are also far shorter than those observed between the basins in *L. cyclurus*.

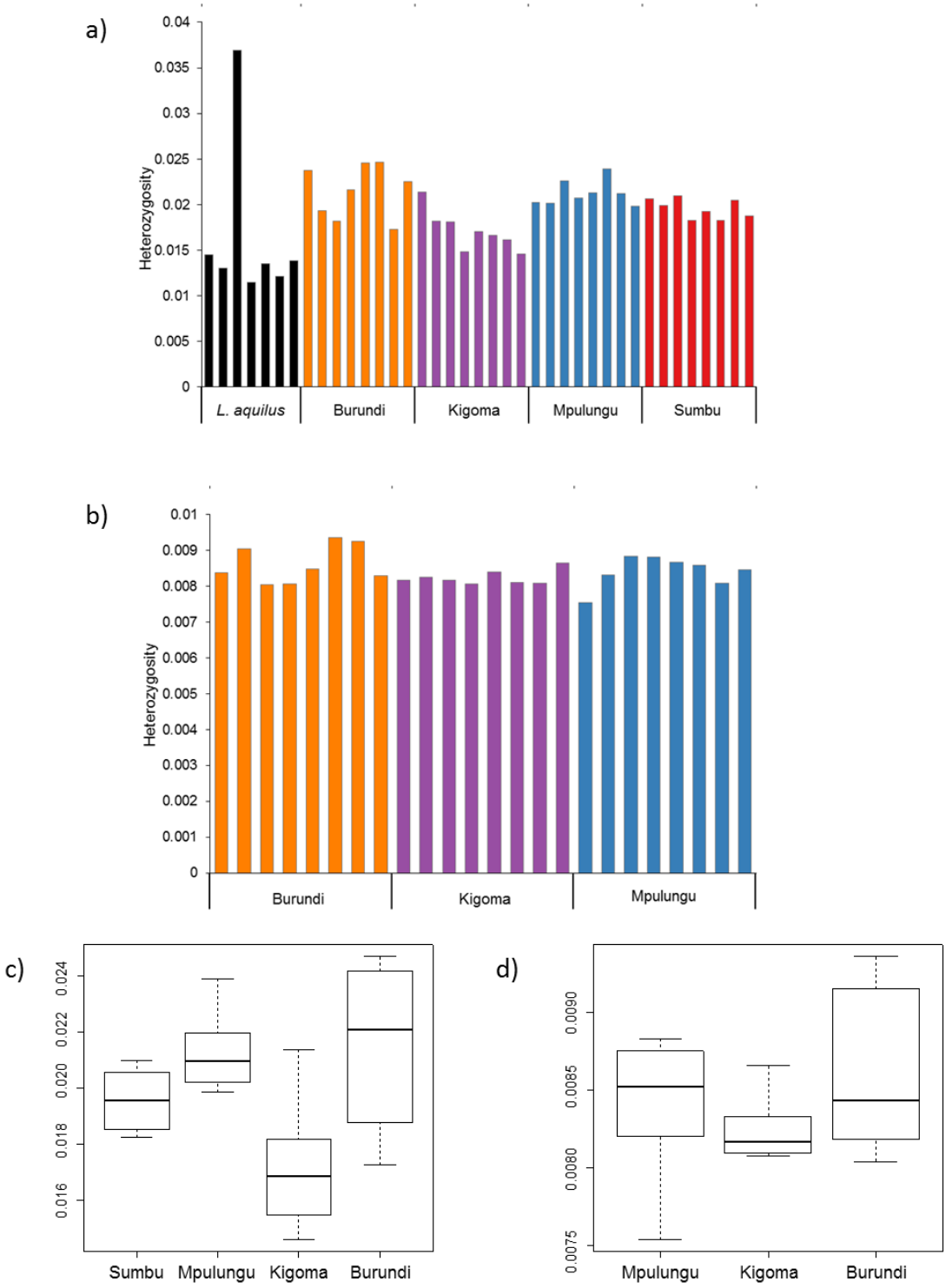


Figure 4 Bar charts show heterozygosity in (a) *Lophiobagrus* and (b) *S. multipunctatus.* Boxplots show heterozygosity in (c) *L. cyclurus* and (d) *S. multipunctatus.*

3.4.3 ABBA/BABA SNPs

The ABBA/BABA test on *L. cyclurus* yielded the following values of the D-statistic, Tree 1: (((Burundi, Kigoma), Mpulungu), *L. aquilus*), D-statistic = 0.0496, without C228 D-statistic = 0.0673. Tree 2: (((Burundi, Kigoma), Sumbu), *L. aquilus*), D-statistic = 0.0569, without C228 D-statistic = 0.0736. Tree 3: (((Mpulungu, Sumbu), Burundi), *L. aquilus*), D-statistic= 0.0110, without C228 D-statistic = 0.0556. Tree 4: (((Mpulungu, Sumbu), Kigoma), *L. aquilus*), D-statistic = 0.0188, without C228 D-statistic = 0.0628 (Figure 5, without C228 Appendix 2, Figure 6). Subsampling of the data at 99%, 95%, 90%, 80% and 70% yielded similar values of the D-statistic with standard errors that do not include zero (Figure 5, without C228 Appendix 2, Figure 6).

3.4.4 Geographic signature extends to local scales

Genetic distances (p-distance) in *L. cyclurus* increased significantly with straight line distances (Mantel test r=0.762, p=0.004) and with lake shore distance (Mantel test r=0.759, p=0.001). This significant relationship remained when only the *L. cyclurus* samples from the two populations in Zambia were considered using straight line distances (Mantel test r=0.482, p=0.041) with a lesser relationship using lake shore distance (Mantel test r=0.479, p=0.095). In the *S. multipunctatus* dataset no relationship was found between genetic distance and geographic distance (Mantel test r= 0.01, p=0.456).

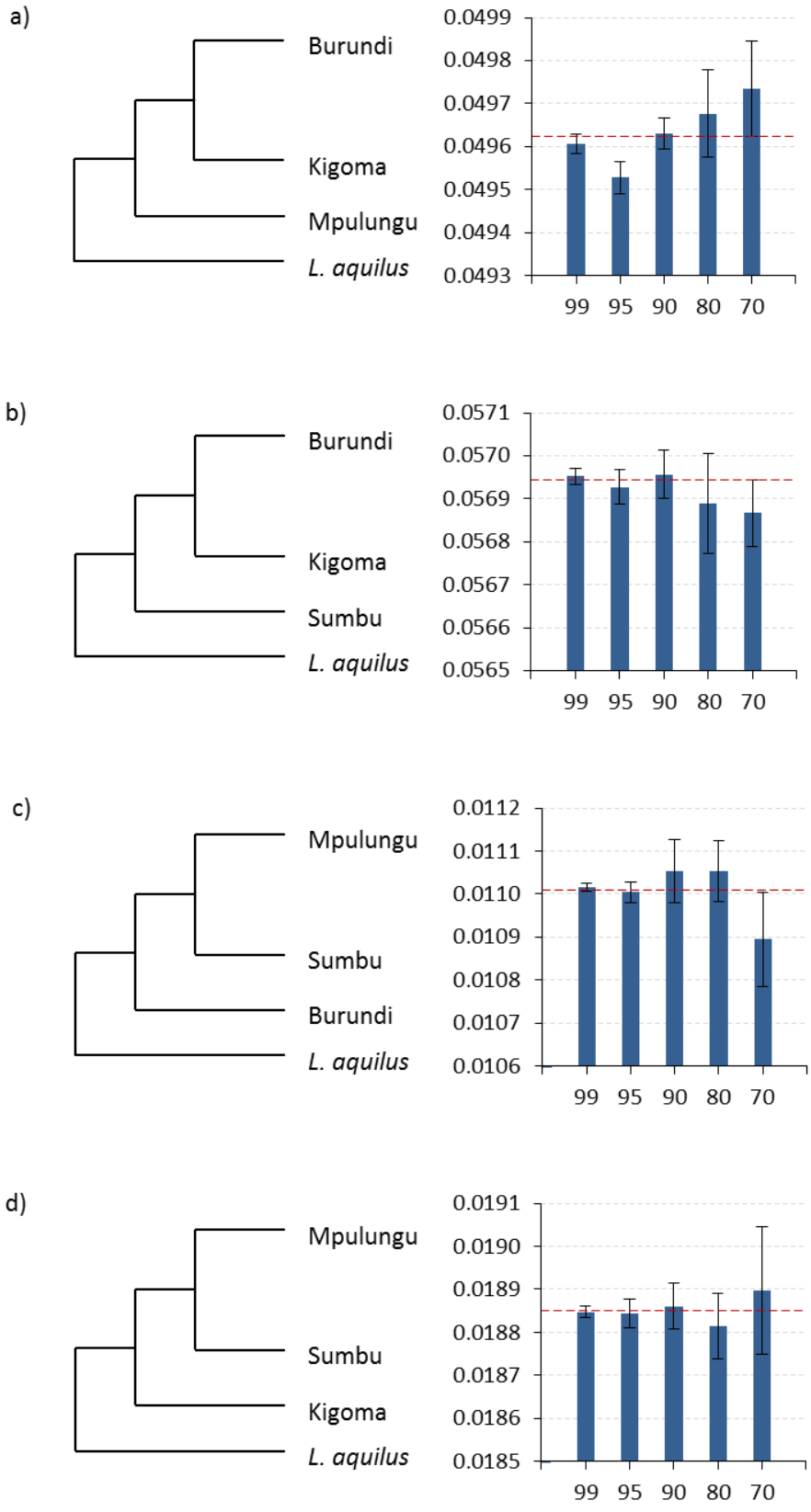


Figure 5 Nominal D Statistics for ABBA-BABA test for *Lophiobagrus.* Red line is the overall nominal D value for each topology. Blue bars represent the mean nominal D value from 1,000 random subsamples of the dataset at each percentage coverage. Error bars are the standard deviation (standard error of the mean is not visible at this scale).

3.5 Discussion

3.5.1 Large scale geographic patterns

There is evidence of different patterns of phylogeographic structure in the two datasets with much stronger phylogeographic structure in *L. cyclurus* than in *S. multipunctatus*. Fossil calibrated molecular dating indicates that both *L. cyclurus* and *S. multipunctatus* diverged from their sister species before the onset of climatic fluctuations in the Late Pleistocene (Chapter 2; Day et al., 2013; Peart et al., 2014), suggesting that the present day geographic patterns of populations in these species do not reflect the location of a recent origin. Rather this distribution is more likely to be a result of their present dispersal ability and perhaps contains an abiding signature of a geographic distribution caused by historic lake level fluctuations.

In *S. multipunctatus* phylogeographic structure along the length of LT is weak. In the ML tree branch lengths are short (Figure 2) but the node separating the populations in the northern basin (Kigoma and Burundi) and southern basin (Mpulungu) is well supported (100%), providing some indication of genetic divergence between these populations. However, this divergence is not reflected in the Fst values (Table 1). Additionally, the PCA analysis resolves divergence between the populations within the northern basin, which are separated by 144km, however this pattern is only observed in this analysis and there is no support for these geographic differences in the ML tree or structure plots.

The *S. multipunctatus* specimens used in this study were sampled from depths up to 100m and this species has previously been sampled from depths up to 170m (Coulter, 1991). The only other species from Lake Tanganyika found at equivalent depths (up to 150m, Coulter, 1991) for which population structure has been investigated is the cichlid *Boulengerochromis microlepis* (Koblmüller et al., 2014) that showed no evidence of lake wide phylogeographic structure based on a 358 bp alignment of the mitochondrial control region. A lack of phylogeographic structure was also observed in benthopelagic cichlid species from Lake Malawi using nuclear microsatellite data including *Rhamphochromis longiceps* (Genner et al., 2008) and multiple *Diplotaxodon* species (Genner et al., 2010; Shaw et al., 2000) though in other *Diplotaxodon* species there was some evidence of restricted gene flow between breeding grounds (Genner et al., 2010).

Synodontis multipunctatus is a brood parasite of multiple species of mouth brooding cichlids (Sato, 1986). Its hosts include both stenotypic species, such as *Tropheus moorii* (Koblmüller et al., 2011; Sefc et al., 2007), which is restricted to rocky shore environments and less geographically restricted species such as *Simochromis diagramma,* which despite also being found in rocky shore environments shows gene flow across sandy areas (Wagner and McCune, 2009) and little phylogeographic structure across large distances (Meyer et al., 1996). Taken together, the broad host range, short branch lengths and lack of phylogeographic structure indicate no evidence that *S. multipunctatus* is becoming specialized to particular hosts including those that are geographically restricted. The wide range of host species with differing ecologies may be partly responsible for the lack of genetic structure in *S. multipunctatus* and the absence of any parental care in this species would also allow the parents to quickly return to deeper waters facilitating long distance dispersal.

In comparison to *S. multipunctatus* strong lake wide phylogeographic structure was shown in all analyses of the *L. cyclurus* data, with the northern basin populations more closely related to each other than to the taxa from the southern basin. The name *Lophiobagrus cyclurus* comes from the synonomisation of *Chrysichthys cyclurus* (Worthington and Ricardo, 1936) and *Lophiobagrus lestradei* (Poll, 1942), which were described from opposite ends of LT. The distributions of many of the catfish species from LT are poorly understood and in recent years, with increased sampling, studies have shown that LT species thought to have large, lake-wide distributions include cryptic diversity (Brown et al., 2010; Chapter 2; Day and Wilkinson, 2006; Peart et al., 2014) and species have been described from single localities (Brown et al., 2011; Wright and Bailey, 2012; Wright and Page, 2006).

The clear genetic differentiation within *L. cyclurus* observed here may be a result of sparse geographic sampling but given the potential for allopatric speciation within LT, which is suggested to be important in the closely related genus *Phyllonemus* (Chapter 2; Peart et al., 2014) the validity of *L. cyclurus* throughout LT should be investigated further. Preliminary analyses were performed on SNPs with no missing data, using species delimitation in SNAPP (Bryant et al., 2012; Leaché et al., 2014), however node heights did not converge after 1,000,000 generations which made path sampling analyses too

computationally expensive to be completed for this study. Regardless of the species status of the different populations they are sufficiently isolated that one future direction could involve reconstructing the independent population history from each location to yield information about demographic history and periods of population growth, an approach that has yielded interesting results in LT cichlid fishes (Koblmüller et al., 2011; Nevado et al., 2013).

Lophiobagrus cyclurus is known to be a paternal mouth brooder (though bi-parental mouth brooding cannot be ruled out due to low sample numbers) (Ochi et al., 2002). This, coupled with its rocky shore littoral habitat, could make it an excellent system to compare to other mouth brooding species of LT catfish such as the bi-parental mouth brooders *Phyllonemus filinemus* (Ochi et al., 2001) and *Phyllonemus typus* (Ochi et al., 2000), and also to LT cichlids that display this behaviour. Population structure has not been investigated in these *Phyllonemus* species, but a previous study indicates that they may be more geographically restricted than previously thought (Chapter 2; Peart et al., 2014), and therefore have the potential for similar patterns of restricted gene flow. Additionally, future sampling from the western shore of LT, a gap currently caused by the ongoing political instability in the Democratic Republic of Congo, would enable the effect of reduced lake levels creating extended suitable shallow habitats across LT to be tested in *L. cyclurus* and compared to patterns observed in the Eretmodini cichlids (Rüber et al., 1999; Verheyen et al., 1996). Colonisation by *L. cyclurus* of the eastern Kigoma site from the western shore, via shallower areas to the north or south, during periods of low lake levels, followed by its subsequent isolation when lake levels rose, could explain the lower heterozygosity value from this site. More data from additional eastern localities would be needed in order to test this hypothesis.

3.5.2 Divergence at local scales

There is evidence for a pattern of isolation by distance in *L. cyclurus* from Zambia between sites separated by 78km (Mantel tests straight line distance p=0.041, lake shore distance p=0.095). The populations from Sumbu and Mpulungu formed well-supported clades in the maximum-likelihood tree (100%, Figure 2), with the exception of a single individual from Sumbu that resolves with the population from Mpulungu (C171). The Fst values also show some divergence (Table 1), although

they are an order of magnitude lower than those between more isolated populations. The STRUCTURE analysis also finds multiple clusters within the Zambian populations depending upon the method used to select K. At K=5 (Figure 2), although the two Zambian populations do not resolve into separate clusters, the Mpulungu samples and C171 are shown to contain the signature of two clusters, one shared with the Sumbu samples, and one predominantly found in these samples. This result holds in the absence of *L. aquilus* (at K=4, Appendix 2, Figure 3). These results are evidence that there is at least some genetic differentiation across small geographic scales in *L. cyclurus*, a pattern seen in multiple cichlid species (e.g., Duftner et al., 2006; Nevado et al., 2013; Sefc et al., 2007; Taylor et al., 2001; Van Steenberge et al., 2013; Wagner and McCune, 2009).

While these results show some evidence of divergence between the populations from Sumbu and Mpulungu, they do not provide any direct information as to what constitutes a barrier to dispersal in this species. The in-flow of the Lufubu river separates these two sampling locations (Figure 1) and previous studies have shown that river inflows with their associated deltas and deposition of sand and mud can constitute barriers to gene flow in rocky shore cichlids (e.g., *Petrochromis* species, Wagner and McCune, 2009). In addition, the fine genetic structure of several cichlid species has been well–studied along the Zambian shore, in which sandy or muddy bays have been shown to be significant barriers to gene flow in some species using mitochondrial and microsatellite loci (e.g., Koblmüller et al., 2011; Sefc et al., 2007). The isolation caused by these barriers has been dated and found to be old in lake terms, suggesting that the barriers are maintained over long time scales allowing for allopatric divergence at small spatial scales. For example, the split between lineages on either side of the muddy Mbete bay in Zambia has been estimated based on mitochondrial data to be between 700,000-945,000 years ago (Sefc et al., 2007). Microallopatry has been suggested as an important mechanism in generating diversity in cichlid fishes in other African great lakes (e.g., Lake Malawi, Fryer, 1959; Rico and Turner, 2002; Van Oppen et al., 1997) leading to the prediction that lineages showing increased genetic structure should also show higher species diversity, but this pattern is not ubiquitous in LT (Taylor et al., 2001). The non-cichlid endemic radiations in LT are much smaller and finer scale investigations into the population structure of *L. cyclurus* would enable tests as to whether microallopatry (as opposed to larger

phylogeographic patterns and other adaptive divergence) played a role in the generation and maintenance of diversity in these taxa too.

Lake-level fluctuations have also been implicated in increasing diversity in cichlids as a result of allopatric divergence followed by secondary admixture when lake levels fall reuniting taxa from previously separated littoral areas, leading to greater divergence in shallow lake shores more vulnerable to lake level fluctuations (Nevado et al., 2013). This has been investigated in cichlid fishes through the reconstruction of population histories based on mitochondrial rates of evolution (Koblmüller et al., 2011; Nevado et al., 2013). However, this dataset featuring many small tags from throughout the genome does not lend itself to investigations of population demographics through coalescent methods. In addition, the large confidence intervals on divergence estimates (from fossil calibration of the phylogeny) do not provide enough temporal resolution to identify the effects of lake level change.

3.5.3 Gene flow in *Lophiobagrus*

In Chapter 2 (Peart et al., 2014) the relationship between *L. aquilus* and different populations of *L. cyclurus* from multiple locations was not resolved in a multilocus nuclear and mitochondrial Bayesian analysis. *Lophiobagrus aquilus* was, however, well supported (1 Bayesian posterior probability) as sister to *L. cyclurus* in a BEAST analysis using only nuclear data, although this relationship was not supported in a *BEAST analysis. The long branch lengths observed in *L. aquilus* in this study compared to those observed between the different *L. cyclurus* populations lends support to the relationship of *L. aquilus* as sister to *L. cyclurus.* This relationship is also supported by morphological characters (Bailey and Stewart, 1984). The genetic isolation of these species is, however, unclear as demonstrated in one specimen, C228 in this study. This specimen, identified as *L. aquilus,* does not resolve in the same clade as other *L. aquilus* specimens in the ML tree, has much higher heterozygosity than any other specimen and in the STRUCTURE analysis is split equally between the cluster for *L. aquilus* and the *L. cyclurus* cluster which is most common in Sumbu (Figure 2). While few inferences can be made from a single individual, this does suggest that there may be some interbreeding between these species; a finding that warrants further investigation. It is also worth noting that specimen C228, was collected in Mpulungu, not Sumbu,

despite its assignment to a genetic cluster most common in specimens from the Sumbu *L. cyclurus* population. The dispersal ability of *L. aquilus* is unknown but it shares the same littoral habitat as *L. cyclurus* and is a paternal mouth brooder (Ochi et al., 2002) so its dispersal potential may be similar to that of *L. cyclurus*.

The results of the ABBA-BABA test suggest that there is some admixture between all of the *L. cyclurus* populations and the overall patterns reflect geographic structure. The values of the D-statistic suggest a greater amount of admixture between the Zambian populations and the closer Kigoma population, despite the low heterozygosity at this site, than is seen between the Zambian populations and the population from Burundi. In order to investigate differential connectivity in different regions of LT further samples would be required, particularly from the western shore. Population subdivision in ancestral species can also influence gene trees and the relative proportion of ABBA/BABA SNPs (Eriksson and Manica, 2012; Slatkin and Pollack, 2008). It is possible that this happened in this dataset due to long standing subdivision between populations in different basins and additional sampling localities would help to reveal this pattern. This may also be the case with the outgroup for these tests, *L. aquilus*, which was sampled only from Zambia in this study but was described from both Zambia and Burundi (Bailey and Stewart, 1984).

The possible shared heritage between the *L. aquilus* sample C228 and *L. cyclurus* is a potential confounding factor, however, when this sample was removed there was an increase in the D-statistic in trees 1 and 2 and a decrease in the value for trees 3 and 4 (Appendix 2, Figure 6) but the overall pattern did not change.

It should be noted that as the dataset could not be aligned to a reference genome it is not possible to formally assess the significance of the D-statistic in these tests by jack-knifing over linkage groups. However, subsampling of the dataset at different levels yielded consistent results, indicating that the results of the D-statistic are robust.

3.5.4 RAD data for studies of recent divergence

Previous studies of allopatric structure in Lake Tanganyika have mainly used short sequences of mitochondrial DNA and microsatellite data (e.g., Koblmüller et al., 2009;2014; Taylor et al., 2001; Wagner and McCune, 2009) with some using AFLPs

(Egger et al., 2007; Mattersdorfer et al., 2012). Mitochondrial data has some advantages, for example, the ability to reconstruct demographic histories through the use of Bayesian skyline plots using established values for the rate of mitochondrial evolution (Koblmüller et al., 2014; Nevado et al., 2013). Disadvantages of mitochondrial data however, include problems of nuclear-mitochondrial discordance, such as mitochondrial sequences being fixed in disparate lineages, and the inability to detect hybridisation due to mono-parental inheritance. In addition, the possibility for sex-biased dispersal means that the signature of dispersal can be masked. Microsatellites are not comparable between taxa, require large sample sizes and provide few datapoints with which to study population divergence. AFLPs provide sites from across the genome but are not comparable between studies and only show dominant markers. RAD data, which provides a very large number of genome-wide sites, has several advantages, and has been used to resolve phylogenies in the rapid Lake Victoria radiation (Wagner et al., 2013) and in phylogeographic and population genetic studies (e.g., Catchen et al., 2013; Emerson et al., 2010; Hohenlohe et al., 2010). There are limitations to the use of RAD data in organisms without a reference genome to align to, for example once RAD sequences are aligned it is possible to investigate areas of differential divergence in the genome (e.g., Hohenlohe et al., 2010; Martin et al., 2013; Nadeau et al., 2013 but see Cruickshank and Hahn, 2014), as well as identifying linkage patterns between areas of the genome. When more genomic resources become available in the future (i.e. a reference genome to align the dataset to) the currently collected data is likely to yield additional information on the focal group.

3.5.5 Conclusions

This study represents the first study of phylogeographic structure in non-cichlid LT fish. *Synodontis multipunctatus* that were sampled from a variety of depths up to 100m showed a low level of phylogeographic structure in contrast to the strong phylogeographic structure seen in *L. cyclurus* sampled from rocky shores. There was also evidence for structure in *L. cyclurus* at smaller spatial scales, across 78km in Zambia, suggesting a role for allopatry in increasing genetic diversity in this taxon.

3.6 References

Bailey, R. M., & Stewart, D. J. (1984) Bagrid catfishes from Lake Tanganyika, with a key and descriptions of new taxa. *Miscellaneous Publications Museum of Zoology, University of Michigan*, **168**

Baird, N. A., Etter, P. D., Atwood, T. S., Currey, M. C., Shiver, A. L., Lewis, Z. A., Selker, E. U., Cresko, W. A., & Johnson, E. A. (2008) Rapid SNP Discovery and Genetic Mapping Using Sequenced RAD Markers. *PloS One*, **3**(10), e3376

Baric, S., Salzburger, W., & Sturmbauer, C. (2003) Phylogeography and evolution of the Tanganyikan cichlid genus Tropheus based upon mitochondrial DNA sequences. *Journal of Molecular Evolution*, **56**(1), 54–68

Brown, K. J., Britz, R., Bills, R., Rüber, L., & Day, J. J. (2011) Pectoral fin loss in the Mastacembelidae: a new species from Lake Tanganyika. *Journal of Zoology*, **284**(4), 286–293

Brown, K. J., Rüber, L., Bills, R., & Day, J. J. (2010) Mastacembelid eels support Lake Tanganyika as an evolutionary hotspot of diversification. *BMC Evolutionary Biology*, **10**, 188

Bryant, D., Bouckaert, R., Felsenstein, J., Rosenberg, N. A., & RoyChoudhury, A. (2012) Inferring species trees directly from biallelic genetic markers: bypassing gene trees in a full coalescent analysis. *Molecular Biology and Evolution*, **29**(8), 1917–1932

Catchen, J., Bassham, S., Wilson, T., Currey, M., O'Brien, C., Yeates, Q., & Cresko, W. A. (2013) The population structure and recent colonization history of Oregon threespine stickleback determined using restriction-site associated DNA-sequencing. *Molecular Ecology*, **22**(11), 2864–2883

Catchen, J. M., Amores, A., Hohenlohe, P., Cresko, W., & Postlethwait, J. H. (2011) Stacks: building and genotyping Loci de novo from short-read sequences. *G3*, **1**(3), 171–182

Chessel, D., Dufour, A., & Thioulouse, J. (2004) The ade4 package-I-One-table methods. *R News*, **4**(June), 5–10

Cohen, A. S., Soreghan, M. J., & Scholz, C. A. (1993) Estimating the age of formation of lakes: An example from Lake Tanganyika, East African Rift system. *Geology*, **21**(6), 511

Coulter, G. W. (1991) The benthic fish communities, In G. W. Coulter (Ed.), *Lake Tanganyika and its Life.* Oxford: Oxford University Press. pp. 151–199

Cruickshank, T. E., & Hahn, M. W. (2014) Reanalysis suggests that genomic islands of speciation are due to reduced diversity, not reduced gene flow. *Molecular Ecology*, **23**(13), 3133–3157

Day, J. J., Peart, C. R., Brown, K. J., Friel, J. P., Bills, R., & Moritz, T. (2013) Continental diversification of an African catfish radiation (Mochokidae: Synodontis) *Systematic Biology*, **62**(3), 351-365

Day, J. J., & Wilkinson, M. (2006) On the origin of the Synodontis catfish species flock from Lake Tanganyika. *Biology Letters*, **2**(4), 548–552

De Mita, S., & Siol, M. (2012) EggLib: processing, analysis and simulation tools for population genetics and genomics. *BMC Genetics*, **13**(1), 27

DePristo, M. A., Banks, E., Poplin, R., Garimella, K. V, Maguire, J. R., Hartl, C., & Daly, M. J. (2011) A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nature Genetics*, **43**(5), 491–498

Dray, S., & Dufour, A. (2007) The ade4 Package: Implementing the duality diagram for ecologists. *Journal of Statistical Software*, **22**(4)

Duftner, N., Sefc, K. M., Koblmüller, S., Nevado, B., Verheyen, E., Phiri, H., & Sturmbauer, C. (2006) Distinct population structure in a phenotypically homogeneous rock-dwelling cichlid fish from Lake Tanganyika. *Molecular Ecology*, **15**(9), 2381–2395

Durand, E. Y., Patterson, N., Reich, D., & Slatkin, M. (2011) Testing for ancient admixture between closely related populations. *Molecular Biology and Evolution*, **28**(8), 2239–2252

Earl, D. A., & vonHoldt, B. M. (2011) STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. *Conservation Genetics Resources*, **4**(2), 359–361

Egger, B., Koblmüller, S., Sturmbauer, C., & Sefc, K. M. (2007) Nuclear and mitochondrial data reveal different evolutionary processes in the Lake Tanganyika cichlid genus Tropheus. *BMC Evolutionary Biology*, **7**, 137

Emerson, K., Merz, C., & Catchen, J. (2010) Resolving postglacial phylogeography using high-throughput sequencing. *Proceedings of the National Academy of Sciences*, **107**(37), 16196–16200

Eriksson, A., & Manica, A. (2012) Effect of ancient population structure on the degree of polymorphism shared between modern human populations and ancient hominins. *Proceedings of the National Academy of Sciences*, **109**(35), 13956–13960

Evanno, G., Regnaut, S., & Goudet, J. (2005) Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. *Molecular Ecology*, **14**(8), 2611–2620

Falush, D., Stephens, M., & Pritchard, J. K. (2003) Inference of population structure using multilocus genotype data: linked loci and correlated allele frequencies. *Genetics*, **164**(4), 1567–1587.

Fryer, G. (1959) The trophic interrelationships and ecology of some of the littoral communities of Lake Nyasa with special reference to the fishes, and a discussion on the evolution of a group of rock-frequenting Cichlidae. *Proceedings of the Zoological Society of London*, **132**, 153–281.

Genner, M. J., Nichols, P., Shaw, P. W., Carvalho, G. R., Robinson, R. L., & Turner, G. F. (2008) Genetic homogeneity among breeding grounds and nursery areas of an exploited Lake Malawi cichlid fish. *Freshwater Biology*, **53**(9), 1823–1831

Genner, M. J., Nichols, P., Shaw, P. W., Carvalho, G. R., Robinson, R. L., & Turner, G. F. (2010) Population structure on breeding grounds of Lake Malawi's "twilight zone" cichlid fishes. *Journal of Biogeography*, **37**(2), 258–269

Genner, M. J., & Turner, G. F. (2012) Ancient hybridization and phenotypic novelty within Lake Malawi's cichlid fish radiation. *Molecular Biology and Evolution*, **29**(1), 195–206

Genner, M. J., Turner, G. F., Barker, S., & Hawkins, S. J. (1999) Niche segregation among Lake Malawi cichlid fishes? Evidence from stable isotope signatures. *Ecology Letters*, **2**, 185–190

Green, R. E., Krause, J., Briggs, A. W., Maricic, T., Stenzel, U., Kircher, M., & Pääbo, S. (2010) A draft sequence of the Neandertal genome. *Science,* **328**(5979), 710–722

Hoffman, J. I., Simpson, F., David, P., Rijks, J. M., Kuiken, T., Thorne, M. a S., & Dasmahapatra, K. K. (2014) High-throughput sequencing reveals inbreeding depression in a natural population. *Proceedings of the National Academy of Sciences*, **111**(10), 3775–3780

Hohenlohe, P. A., Amish, S. J., Catchen, J. M., Allendorf, F. W., & Luikart, G. (2011) Next-generation RAD sequencing identifies thousands of SNPs for assessing hybridization between rainbow and westslope cutthroat trout. *Molecular Ecology Resources*, **11**(Suppl. 1), 117–122

Hohenlohe, P. A., Bassham, S., Etter, P. D., Stiffler, N., Johnson, E. A., & Cresko, W. A. (2010) Population genomics of parallel adaptation in threespine stickleback using sequenced RAD tags. *PLoS Genetics*, **6**(2), e1000862

Huateng, H., & Knowles, L. L. (2014) Excluding missing data from next-generation sequences. *Systematic Biology*, doi:10.1093/sysbio/syu046

Hudson, R. R., Slatkin, M., & Maddison, W. P. (1992) Estimation of levels of gene flow from DNA sequence data. *Genetics*, **132**, 583-589

Jakobsson, M., & Rosenberg, N. A. (2007) CLUMPP: a cluster matching and permutation program for dealing with label switching and multimodality in analysis of population structure. *Bioinformatics*, **23**(14), 1801–1806

Jombart, T., Devillard, S., Dufour, A-B., & Pontier, D. (2008) Revealing cryptic spatial patterns in genetic variability by a new multivariate method. *Heredity*, **101**(1), 92–103

Keller, I., Wagner, C. E., Greuter, L., Mwaiko, S., Selz, O. M., Sivasundar, A., & Seehausen, O. (2013) Population genomic signatures of divergent adaptation, gene flow and hybrid speciation in the rapid radiation of Lake Victoria cichlid fishes. *Molecular Ecology*, **22**(11), 2848–2863

Koblmüller, S., Duftner, N., Sefc, K. M., Aigner, U., Rogetzer, M., & Sturmbauer, C. (2009) Phylogeographic structure and gene flow in the scale-eating cichlid Perissodus microlepis (Teleostei, Perciformes, Cichlidae) in southern Lake Tanganyika. *Zoologica Scripta*, **38**(3), 257–268

Koblmüller, S., Odhiambo, E. a., Sinyinza, D., Sturmbauer, C., & Sefc, K. M. (2014) Big fish, little divergence: phylogeography of Lake Tanganyika's giant cichlid, Boulengerochromis microlepis. *Hydrobiologia,* doi:10.1007/s10750-014-1863-z

Koblmüller, S., Salzburger, W., Obermüller, B., Eigner, E., Sturmbauer, C., & Sefc, K. M. (2011) Separated by sand, fused by dropping water: habitat barriers and fluctuating water levels steer the evolution of rock-dwelling cichlid populations in Lake Tanganyika. *Molecular Ecology*, **20**(11), 2272–2290

Koblmüller, S., Sturmbauer, C., Verheyen, E., Meyer, A., & Salzburger, W. (2006) Mitochondrial phylogeny and phylogeography of East African squeaker catfishes (Siluriformes: Synodontis) *BMC Evolutionary Biology*, **6**, 49

Leaché, A. D., Fujita, M. K., Minin, V. N., & Bouckaert, R. R. (2014) Species Delimitation using Genome-Wide SNP Data. *Systematic Biology,* doi:10.1093/sysbio/syu018

Marijnissen, S. A. E., Michel, E., Kamermans, M., Olaya-Bosch, K., Kars, M., Cleary, D. F. R., & Menken, S. B. J. (2008) Ecological correlates of species differences in the Lake Tanganyika crab radiation. *Hydrobiologia*, **615**(1), 81–94

Martin, S. H., Dasmahapatra, K. K., Nadeau, N. J., Salazar, C., Walters, J. R., Simpson, F., & Jiggins, C. D. (2013) Genome-wide evidence for speciation with gene flow in Heliconius butterflies. *Genome Research*, **23**(11), 1817–1828

Mattersdorfer, K., Koblmüller, S., & Sefc, K. M. (2012) AFLP genome scans suggest divergent selection on colour patterning in allopatric colour morphs of a cichlid fish. *Molecular Ecology*, **21**(14), 3531–3544

McGlue, M. M., Lezzar, K. E., Cohen, A. S., Russell, J. M., Tiercelin, J.-J., Felton, A. A., & Nkotagu, H. H. (2008) Seismic records of late Pleistocene aridity in Lake Tanganyika, tropical East Africa. *Journal of Paleolimnology*, **40**(2), 635–653

McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., & DePristo, M. A. (2010) The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Research*, **20**, 1297–1303

Meyer, A., Knowles, L. L., & Verheyen, E. (1996) Widespread geographical distribution of mitochondrial haplotypes in rock-dwelling cichlid fishes from Lake Tanganyika. *Molecular Ecology*, **5**(3), 341–350

Nadeau, N. J., Martin, S. H., Kozak, K. M., Salazar, C., Dasmahapatra, K. K., Davey, J. W., & Jiggins, C. D. (2013) Genome-wide patterns of divergence and gene flow across a butterfly radiation. *Molecular Ecology*, **22**(3), 814–826

Nevado, B., Mautner, S., Sturmbauer, C., & Verheyen, E. (2013) Water-level fluctuations and metapopulation dynamics as drivers of genetic diversity in populations of three Tanganyikan cichlid fish species. *Molecular Ecology*, **22**(15), 3933–3948

Ntakimazi, G. (2006a) Lophiobagrus cyclurus. *The IUCN Red List of Threatened Species. Version 2014.2*. <www.iucnredlist.org>. Downloaded on 12th September 2014

Ntakimazi, G. (2006b) Synodontis multipunctatus. *The IUCN Red List of Threatened Species. Version 2014.2.* <www.iucnredlist.org>. Downloaded on 12th September 2014

Ochi, H., Rossiter, A., & Yanagisawa. (2000) The first record of a biparental mouthbrooding catfish. *Journal of Fish Biology*, **57**(6), 1601–1604

Ochi, H., Rossiter, A., & Yanagisawa, Y. (2001) Biparental mouthbrooding of the catfish Phyllonemus filinemus in Lake Tanganyika. *Ichthyological Research*, **48**, 225–229

Ochi, H., Rossiter, A., & Yanagisawa, Y. (2002) Paternal mouthbrooding bagrid catfishes in Lake Tanganyika. *Ichthyological Research*, **49**, 270–273

Paradis, E., Claude, J., & Strimmer, K. (2004) APE: Analyses of Phylogenetics and Evolution in R language. *Bioinformatics*, **20**(2), 289–290

Peart, C. R., Bills, R., Wilkinson, M., & Day, J. J. (2014) Nocturnal claroteine catfishes reveal dual colonisation but a single radiation in Lake Tanganyika. *Molecular Phylogenetics and Evolution*, **73**, 119–128

Poll, M. (1942) Description d'un genre nouveau de Bagridae du lac Tanganika. *Revue de Zoologie et de Botanique Africaines*, **35**(3), 318–322

Pritchard, J. K., Stephens, M., & Donnelly, P. (2000) Inference of population structure using multilocus genotype data. *Genetics*, **155**(2), 945–959

Rabosky, D. L., Santini, F., Eastman, J., Smith, S. A, Sidlauskas, B., Chang, J., & Alfaro, M. E. (2013) Rates of speciation and morphological evolution are correlated across the largest vertebrate radiation. *Nature Communications*, **4**, 1958

Reich, D., Thangaraj, K., Patterson, N., Price, A. L., & Singh, L. (2009) Reconstructing Indian population history. *Nature*, **461**(7263), 489–494

Rheindt, F. E., Fujita, M. K., Wilton, P. R., & Edwards, S. V. (2014) Introgression and phenotypic assimilation in Zimmerius flycatchers (Tyrannidae): Population

genetic and phylogenetic inferences from genome-wide SNPs. *Systematic Biology*, **63**(2), 134-152

Rico, C., & Turner, G. F. (2002) Extreme microallopatric divergence in a cichlid species from Lake Malawi. *Molecular Ecology*, **11**(8), 1585–1590

Robertson, D. R., & Smith-Vaniz, W. F. (2008) Rotenone: an essential but demonized tool for assessing marine fish diversity. *BioScience*, **58**(2), 165–170

Rosenberg, N. A. (2004) Distruct: a program for the graphical display of population structure. *Molecular Ecology Notes*, **4**(1), 137–138

Rüber, L., Verheyen, E., & Meyer, A. (1999) Replicated evolution of trophic specializations in an endemic cichlid fish lineage from Lake Tanganyika. *Proceedings of the National Academy of Sciences*, **96**(18), 10230–10235.

Salzburger, W., Baric, S., & Sturmbauer, C. (2002) Speciation via introgressive hybridization in East African cichlids? *Molecular Ecology*, **11**(3), 619–625

Sato, T. (1986) A brood parasitic catfish of mouthbrooding cichlid fishes in Lake Tanganyika. *Nature*, **323**, 58–59

Seehausen, O., Terai, Y., Magalhaes, I. S., Carleton, K. L., Mrosso, H. D. J., Miyagi, R., & Okada, N. (2008) Speciation through sensory drive in cichlid fish. *Nature*, **455**(7213), 620–626

Sefc, K. M., Baric, S., Salzburger, W., & Sturmbauer, C. (2007) Species-specific population structure in rock-specialized sympatric cichlid species in Lake Tanganyika, East Africa. *Journal of Molecular Evolution*, **64**(1), 33–49

Shaw, P. W., Turner, G. F., Idid, M. R., Robinson, R. L., & Carvalho, G. R. (2000) Genetic population structure indicates sympatric speciation of Lake Malawi pelagic cichlids. *Proceedings of the Royal Society B: Biological Sciences*, **267**(1459), 2273–2280

Slatkin, M., & Pollack, J. L. (2008) Subdivision in an ancestral species creates asymmetry in gene trees. *Molecular Biology and Evolution*, **25**(10), 2241–2246

Stamatakis, A. (2006) RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics,* **22**(21), 2688–2690

Taylor, M. I., Rüber, L., & Verheyen, E. (2001) Microsatellites reveal high levels of population substructuring in the species-poor Eretmodine cichlid lineage from Lake Tanganyika. *Proceedings of the Royal Society B: Biological Sciences*, **268**(1469), 803–808

Van Oppen, M. J. H., Turner, G. F., Rico, C., Deutsch, J. C., Ibrahim, K. M., Robinson, R. L., & Hewitt, G. M. (1997) Unusually fine-scale genetic structuring found in rapidly speciating Malawi cichlid fishes. *Proceedings of the Royal Society B: Biological Sciences*, **264**(1389), 1803–1812

Van Steenberge, M., Vanhove, M. P. M., Breman, F. C., & Snoeks, J. (2013) Complex geographical variation patterns in Tropheus duboisi Marlier, 1959 (Perciformes, Cichlidae) from Lake Tanganyika. *Hydrobiologia*, doi: 10.1007/s10750-013-1692-5

Verheyen, E., Rüber, L., Snoeks, J., & Meyer, A. (1996) Mitochondrial phylogeography of rock-dwelling cichlid fishes reveals evolutionary influence of historical lake level fluctuations of Lake Tanganyika, Africa. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, **351**(1341), 797–805

Wagner, C. E., Keller, I., Wittwer, S., Selz, O. M., Mwaiko, S., Greuter, L., & Seehausen, O. (2013) Genome-wide RAD sequence data provide unprecedented resolution of species boundaries and relationships in the Lake Victoria cichlid adaptive radiation. *Molecular Ecology*, **22**(3), 787–798

Wagner, C. E., & McCune, A. R. (2009) Contrasting patterns of spatial genetic structure in sympatric rock-dwelling cichlid fishes. *Evolution,* **63**(5), 1312–1326

Willing, E.-M., Dreyer, C., & van Oosterhout, C. (2012) Estimates of genetic differentiation measured by F(ST) do not necessarily require large sample sizes when using many SNP markers. *PloS One*, **7**(8), e42649

Winkelmann, K., Genner, M. J., Takahashi, T., & Rüber, L. (2014) Competition-driven speciation in cichlid fish. *Nature Communications*, **5**, 3412

Worthington, E. B., & Ricardo, O. K. (1936) The fish of Lake Tanganyika (other than Cichlidae) *Proceedings of the Zoological Society of London*, **106**(4), 1061–1112

Wright, J. J., & Bailey, R. M. (2012) Systematic revision of the formerly monotypic genus Tanganikallabes (Siluriformes: Clariidae) *Zoological Journal of the Linnean Society*, **165**(1), 121–142

Wright, J. J., & Page, L. M. (2006) Taxonomic revision of Lake Tanganyikan Synodontis (Siluriformes: Mochokidae) *Bulletin of the Florida Museum of Natural History*, **46**(4), 99–154

Zerbino, D. R., & Birney, E. (2008) Velvet: algorithms for de novo short read assembly using de Bruijn graphs. *Genome Research*, **18**(5), 821–829

Chapter 4 Do sympatric catfish radiations in Lake Tanganyika show patterns of ecological diversification?

4.1 Abstract

The process of adaptive radiation has been proposed to lead to divergent morphologies and ecologies in the East African Rift Lakes but this remains to be investigated in many endemic radiations. Lake Tanganyika is home to a host of endemic radiations, allowing comparative studies investigating generalisations in patterns of diversification to be conducted. There are three independent catfish radiations in Lake Tanganyika. The two largest, in the genus *Synodontis* (≈11 species) and the sub-family Claroteinae (15 species described with additional diversity) are found in sympatry, allowing comparisons of the timing and mode of diversification. In this study, these radiations were placed in a common phylogenetic context for the first time using a two-step dating process with multiple fossil calibrations. This phylogeny allowed the assessment of patterns of morphological and ecological diversification, with stable isotopes used to estimate resource use. The *Synodontis* radiation was dated as later than previous estimates, more concurrent with the individual genera within the claroteine radiation than the radiation as a whole, and coinciding with the onset of lacustrine conditions in the southern basin of Lake Tanganyika. The claroteine radiation is characterised by an early drop in subclade disparity in both the morphological and isotopic analyses, followed by a period of stability, suggesting initial niche filling coinciding with the diversification into different genera, potentially into feeding niches, followed by non-adaptive diversification. In the *Synodontis* radiation disparity is mainly partitioned within subclades and there is evidence of niche conservatism with overlap between clades in both diet and morphology.

4.2 Introduction

The East African Great Lakes are hotspots of biodiversity. This high diversity includes many endemic radiations where multiple species have rapidly evolved

from a common ancestor, including the hyperdiverse cichlid fishes that have diversified into a multitude of different body forms and niche specialisms (Fryer and Iles, 1972). A radiation can be characterised as adaptive under the ecological theory when genetic diversification is accompanied by a change in resource exploitation, with a concomitant change in physiological, behavioural or morphological traits required to exploit these new resources, producing an ecologically diverse array of descendent species (Schluter, 1996; Schluter, 2000). Adaptive radiations may occur as species diversify to fill different environments, as alternate food resources are more prevalent in different environments, or may occur in syntopy as a result of niche partitioning. Not all radiations are necessarily adaptive under this definition however, for example, when genetic diversification is as a result of sexual selection, hybridisation or geographic isolation (e.g., woodland salamanders, Kozak, et al., 2006) little or no divergence along ecological axes is observed, although hybridisation has been suggested to lead to novel traits in some radiations (Genner and Turner, 2012; Keller et al., 2013). Indeed, in some cases a selective pressure to retain ancestral characteristics leads to niche conservatism within radiations (Kozak and Wiens, 2006).

Adaptive radiations would be expected to initially show rapid rates of ecological diversification as new species separate into disparate niches, followed by a transition to slower, equilibrium rates as niche space within these new adaptive zones fills (Freckleton and Harvey, 2006). In contrast, non-adaptive radiations would be expected to conform to Brownian motion, as the same original area of niche space continues to be randomly filled by genetically divergent species.

The cichlid fishes are the most intensively studied of the East African Rift Lake radiations and multiple factors have in implicated in the generation of their diversity in addition to niche partitioning and ecological divergence (e.g., Albertson, 2008), these include intralacustrine allopatry (discussed in Chapter 3, e.g., Koblmüller et al., 2011; Verheyen et al., 1996), hybridisation (Genner and Turner, 2012; Salzburger et al., 2002), sensory drive (Seehausen et al., 2008), and competition (Winkelmann et al., 2014). In cichlids, phenotypes, involving changes in trophic morphology, have evolved repeatedly across the East African Rift Lakes (Cooper et al., 2010; Fryer and Iles, 1972) and can also be convergent within a radiation (e.g., Muschick et al., 2012; Rüber et al., 1999) suggesting that the same

selective pressures may lead to the same evolutionary results. Changes in trophic morphology have also been found to correlate with additional changes in body shape (Rüber and Adams, 2001), a result that was independent of phylogeny. In addition, changes in body shape are correlated with feeding preferences across different LT cichlid tribes (Clabaut et al., 2007). The extent that similar patterns to those observed in the cichlids are found in other East African Rift Lake endemic radiations remains to be elucidated.

Many species co-exist in the littoral zone of the East African Great Lakes but the role of niche partitioning in the maintenance of this diversity is unclear. In LT cichlids there is evidence of niche partitioning (e.g., Sturmbauer et al., 1992), however, niche partitioning can also be incomplete (e.g., overlap between shrimp eating cichlids despite differences in preferred prey size, Yuma et al., 1998) and convergent species have been shown to coexist (Muschick et al., 2012). In addition, facultative commensalisms are also prevalent in cichlids (Hori et al., 1993) that may also facilitate coexistence. Niche partitioning is also observed in *Platythelphusa* crabs (Marijnissen et al., 2008) though the extent that this allows co-existence of multiple species in the littoral zone or has influenced their diversification remains to be tested.

In LT there are three independent catfish radiations, in the genus *Tanganikallabes* (expanded from a monotypic genus to three species, Wright and Bailey, 2012), and two larger radiations that are used in this study, in the genus *Synodontis* (≈ 11 species) (Day and Wilkinson, 2006; Koblmüller et al., 2006) and the subfamily Claroteinae (15 species described with additional diversity) (Chapter 2; Peart et al., 2014). The timing of colonisation for the *Tanganikallabes* radiation has not yet been established, but claroteine and *Synodontis* radiations have been dated as diversifying at similar times in previous independent studies based on single fossil calibrations (claroteine 5.08 Ma, 3.61-6.84 Ma) (Chapter 2; Peart et al., 2014), *Synodontis* 7.9 Ma (5.7-10 Ma) (Day et al., 2013). Dates based on single fossil calibrations can be problematic as they are heavily influenced by the incomplete nature of the fossil record and are very sensitive to incorrect placement of the fossil on the phylogeny or incorrect taxonomic assignment. In the case of the claroteine radiation there are additional issues concerning the placement of the fossil used to date the radiation (discussed in Chapter 2). In *Synodontis* the addition of further taxa from outside LT in the genus *Synodontis* and

a change in the single fossil calibration used caused the date estimate of the LT radiation to change from 5.5 Ma (4.0–7.3 HPD) (Day et al., 2009) to 7.9 Ma (5.7–10 HPD) (Day et al., 2013). It is problematic to assign fossils to the genus *Synodontis* as it is defined using soft tissues characters (Vigliotta, 2008). In order to establish a robust comparison of the timing of diversification between these two radiations this study seeks to place them in a common phylogenetic context.

The adaptive character of these radiations is currently unknown, as is the relative importance of other factors in their diversification, though there is evidence that geographic restriction plays a role in at least two claroteine genera, *Phyllonemus* and *Lophiobagrus* (Chapter 2; Chapter 3). There is no clear sexual dimorphism in either of the catfish radiations but several claroteine taxa are known paternal or bi-parential mouthbrooders (Ochi et al., 2000; 2001; 2002), which may provide a morphological constraint with respect to feeding. Coulter (1991) has suggested that LT *Chrysichthys* (defined in Chapter 2) are partitioned by habitat and depth though there is some overlap between species. Many claroteine species can, however, be collected in sympatry (pers. obs.) and the extent to which resource partitioning was important in their diversification or co-existence is currently unknown.

In the LT *Synodontis* radiation the role of geographic isolation is not yet clear with *S. multipunctatus* showing little genetic structure along the length of LT (Chapter 3), however, multiple *Synodontis* species have been described from single localities (Wright and Page, 2006) and the extent of geographic restriction throughout the radiation is unknown. There is some evidence that LT *Synodontis* species have differing diets (Wright and Page, 2006, and references therein), however, this has not been investigated between sympatric species. The extent of divergence in mating strategies between *Synodontis* species remains to be investigated, though *S. multipunctatus* is known to be a brood parasite of multiple species of mouth brooding cichlids (Sato, 1986). The LT *Synodontis* have been shown to be Müllerian mimics (Wright, 2011) with warning colouration on their fins, however whether this protection is further enhanced by shared behaviours is unknown.

Integrated studies allow powerful inferences into adaptive radiation to be made by using different datasets to address the same questions. This study incorporates multiple linear measurements of morphological features thought to

be ecologically significant, and diet is also measured using stable isotope values. In cichlid fishes studies have tended to use landmark based geometric morphometrics as opposed to linear measurements (e.g., Clabaut et al., 2007; Muschick et al., 2012; Rüber and Adams, 2001), however, 2-dimensional geometric morphometric analyses are difficult to conduct on benthic taxa that have a more depressed shape and ecologically important changes in body width can be missed. In addition, the use of linear measurements between morphological features allows the same features to be compared across taxa with very different body shapes. Because this study is the first to investigate morphological changes across the phylogeny, measurements were taken from the entire body. In contrast, other studies have made inferences about ecomorphological divergence using only parts of the anatomy, for example, opercle shape in notothenioid icefishes (Wilson et al., 2013) and in three-spined sticklebacks (Arif et al., 2009).

This study places both of the independent LT catfish radiations in the same phylogenetic context for the first time using a two-step approach that uses multiple fossil calibrations on a phylogeny of the Ostariophysians (using three nuclear markers) in order to provide an age estimate for the origin of the 'Big Africa' clade (Sullivan et al., 2006) that encompasses the majority of African catfishes. This estimate is then used as a prior in a subsequent analysis of just the 'Big Africa' clade including the best available sampling of both LT catfish radiations using four nuclear markers and two mitochondrial markers. This second phylogeny, in combination with a dataset of ten morphological measurements for each specimen and an estimate of niche for sympatric Zambian species from stable isotopes of carbon and nitrogen, is used to address the following questions; i) Did the two independent catfish radiations diversify concurrently? ii) Is there evidence that these radiations are adaptive? iii) Do the sympatric species in each radiation from the southern basin show any evidence of resource partitioning? iv) Does each radiation follow the same pattern in terms of changing disparity through time?

4.3 Methods

4.3.1 Taxonomic sampling

Specimens from LT, including samples for genetic and isotopic analysis (Zambia only) were collected using a variety of methods as described in Chapter 2 from

sites in Burundi, Tanzania and Zambia. In addition, specimens from museum collections were measured for inclusion in the morphological analysis with a full list given in Appendix 3 (Table 1).

Within the claroteine radiation there is taxonomic uncertainty over the species status of *Chrysichthys grandis* and *Chrysichthys graueri* with conflicting designations in different taxonomic keys (Bailey and Stewart, 1984; Hardman, 2008), and both species were synonymised in the catalogue of the Royal Museum of Central Africa, Tervuren. Following the taxonomy of Bailey and Stewart (1984) specimens with a longer lower jaw are assigned to *Chrysichthys platycephalus* whereas following the key of Hardman (2008) a longer lower jaw is a feature of *C. graueri*. It is worth noting that neither of these studies examined the type specimens of *C. graueri* and Hardman cites no specimens at all. Chapter 2 (Peart et al., 2014) included tissues from one specimen with a longer lower jaw (CU95204) which resolved within *C. platycephalus* (which have jaws of equal length) in an analysis based on nuclear and mitochondrial genes. Therefore in this study specimens with a longer lower jaw are classed as *C. platycephalus*. There are two *Chrysichthys* clades with longer upper jaws in the molecular phylogeny (Chapter 2; Peart et al., 2014) including CU95203 which has been measured. This specimen is accessioned as *Chrysichthys acsiorum* and has the small teeth of this species, however, it does not fit the other measurements used to describe *C. acsiorum*. The ratios of different measurements, however, do fit *C. graueri* as described in Bailey and Stewart. Therefore in this study the key of Bailey and Stewart (1984) is used along with the addition of the more recently described species *C. acsiorum* from Hardman (2008).

In the genus *Phyllonemus* there is additional diversity that has not yet been formally described (Bills, pers. obs.), with five species supported by Bayesian species delimitation methods in Chapter 2 (Peart et al., 2014). Of the two putative species that were not supported, this study excludes *Phyllonemus* sp. A, which had very low support in the species delimitation analysis (range from 0.25-0.32), but includes *Phyllonemus* sp. D (support from 0.8-0.9) in an attempt to include the most recent divergences. In addition, samples of *L. cyclurus* from multiple localities are included (from Burundi, Kigoma in Tanzania and Zambia) as these show genetic divergence in Chapter 3. These taxa are included in order to encompass as much on-going divergence in the radiation as possible.

The taxonomic key for the LT *Synodontis* radiation Wright and Page (2006) was published after the first molecular phylogenies (Day and Wilkinson, 2006; Koblmüller et al., 2006) and described a further three species. This key provides some useful diagnostic features, including for the first time, the axillary pore which has not previously been tested to assess its use as an informative character. However, molecular phylogenies suggest that some of the features in the key are not sufficient to diagnose species. Molecular data from specimens used in the key are rare. However, CU88758 and BMNH 2005.9.26.18[1] are both in the Wright and Page, (2006) key as *S. polli* and are included in a phylogeny constructed using Cytochrome b (*Cytb*) sequences in Appendix 3 (Figure 1). In this phylogeny, these specimens resolve in distinct clades and within the LT clade are not closely related. The key separates *S. petricola* and *S. lucipinnis* by the difference between a very small and absent axillary pore and the absence/presence of light patches at the base of black triangles on rayed fins. In the *Cytb* phylogeny the clade that contained specimens with these light patches also included specimens without them suggesting that this character may not be useful in species diagnostics or may be affected by preservation. In contrast the specimens which key out as *S. petricola* or *S. lucipinnis* cluster by sampling location, in either the northern or southern basin, in a mitochondrial phylogeny, a pattern which has been seen in other LT taxa (Brown et al., 2010; Peart et al., 2014). In addition, white colour spines and papillae shape are also used as a diagnostic features but the colours can fade during preservation and some papillae shapes are delicate and do not preserve well. These difficulties in accurately identifying specimens suggest that the LT *Synodontis* key requires refinement and there is a problem in relating specimens from museum collections to the taxonomy suggested by the molecular phylogeny. Several species are very distinctive, (*S. granulosus*, *S. multipunctatus*) and so can be accurately identified from museum collections, however, for the other species, this study has included only specimens for which *Cytb* sequences were available (32 from GenBank and 106 generated for this study, Appendix 3, Table 2) and investigated clades in this phylogeny rather than named species (Appendix 3, Figure 1). *Cytb* sequences were used as this marker has been found to provide resolution in this genus (Day and Wilkinson, 2006; Day et al., 2009; Koblmüller et

[1] referred to in Wright and Page 2006 as BMNH 2005.9.26.17-18, however BMNH 2005.9.26.17 using the same key is the morphologically distinct *S. irascae*

al., 2006) and there has been no evidence of nuclear-mitochondrial discordance in this group (Day et al., 2009). The only *S. grandiops* specimens for which *Cytb* sequences are available is CU91902 which resolves within *S. multipunctatus*, however, this specimen was not measured for this study so its identity could not be established. Due to this, this specimen was not used to place *S. grandiops* in the molecular phylogeny but this species was still included (using measurements from the type series) in the non-phylogenetically corrected analyses. All of the *S. multipunctatus* specimens used in this study do not conform to the description of the more recently described *S. grandiops* (Wright and Page, 2006).

4.3.2 Phylogenetic tree construction

123 novel DNA sequences were generated in order to build the phylogenies and combined with 199 unpublished sequences generated in the laboratory of Dr Thomas Near and 252 sequences downloaded from GenBank including sequences generated for Chapter 2. A full list is given in Appendix 3 (Table 2). For the DNA sequences that were generated for this study, DNA was extracted using the Qiagen DNeasy kit. PCR reactions were conducted, cleaned, checked and sequenced with the same components and reaction conditions as those in Chapter 2. Primers and annealing temperatures were taken from the literature for RAG1 exon 3 (Sullivan et al., 2006), ENC1 (Li et al., 2007), Plagl2 (Li et al., 2007), RAG2 (Sullivan et al., 2006). Chromatograms were checked and edited in Geneious 5.6 (Biomatters), Sequences were aligned with Geneious align and translated in order to check for stop codons by eye. The sequences were trimmed to reduce missing data and match previously published sequences leading to 1371bp RAG1, 810bp ENC1, 680bp Plagl2, 922bp RAG2, 759bp (1st and 2nd positions *Cytb*) and 434bp (1st and 2nd positions CO1).

In order to compare the timing of diversification events in both the *Synodontis* and claroteine LT radiations, it is necessary to place them in the same dated phylogenetic context. Both of these radiations resolve within a 'Big Africa' clade (Sullivan et al., 2006) containing the majority of siluriform genera found in Africa. The oldest catfish fossil found in Africa, in the genus *Nigerium*, is attributed to the family Claroteidae within the 'Big Africa' clade, however accurate placement is problematic (see Chapter 2 for a discussion regarding the placement of this fossil). There is also a shortage of siluriform fossils that can be placed accurately

on the phylogeny and a rapid radiation of clades at the base of the siluriform phylogeny which often forms an unresolved polytomy (Lundberg et al., 2007; Sullivan et al., 2006) making inferences on the timing of one clade from a fossil placed in another siluriform clade difficult. To account for this a two-step dating procedure was used that involved dating the onset of the 'Big Africa' clade (Sullivan et al., 2006) using a phylogeny including taxa and calibrations from throughout the Ostariophysians. The 'Big Africa' posterior probability estimate was then used as a prior in a 'Big Africa' only analysis using more molecular markers and denser taxonomic sampling, including the two Lake Tanganyika radiations.

The Ostariophysian tree was constructed using 173 specimens from across the Ostariophysians in 146 genera, with two outgroup taxa from the Clupeiformes and five fossil constraints. The following fossil calibrations are taken from Near et al., (2012) and applied with the same priors, a rationale for these priors, a discussion of the formations they are found in and the character states that justify their placements is provided in that paper. The fossil genus †*Rubiesichthys* (Poyato-Ariza, 1996), which resolves in a clade with the genus *Chanos* in an analysis based on morphological characters (Poyato-Ariza et al., 2010), was used as a stem *Chanos* calibration. This calibration was applied as a lognormal prior, mean = 1.51 and SD = 0.8 which gives 133.9 Ma as the minimal age offset and 150.8 Ma as the 95% soft upper bound. In Near et al., (2012) this calibration was used to date the most recent common ancestor (MRCA) of *Chanos* and *Cromeria*. In this analysis there is increased taxon sampling from the Gonorynchiformes with the genera *Chanos*, *Cromeria*, *Grasseichthys*, *Gonorynchus*, *Kneria*, *Parakneria*, and *Phractolaemus* included. There is conflict in the placement of the family Gonorynchidae between molecular (*Gonorynchus* sister to a clade containing *Chanos*, *Cromeria*, *Grasseichthys*, *Parakneria* and *Kneria*) (Lavoué et al., 2005) and morphological analyses (Gonorynchidae and Kneriidae in a clade sister to Chanidae) (Poyato-Ariza et al., 2010). In this analysis the calibration prior was applied to the MRCA *Chanos*, *Cromeria*, *Grasseichthys*, *Kneria*, *Parakneria*, and *Phractolaemus* with no monophyly constraint. This allows the calibration to represent a stem *Chanos* lineage in both topology hypotheses. The fossil †*Astephus* (Lundberg, 1975) which resolves as sister to the Ictaluridae in phylogenetic analyses based on morphological characters (Lundberg, 1992) is used to date the

MRCA of Ictaluridae (*Ameiurus, Ictalurus, Noturus, Pylodictis*) in this analysis and *Cranoglanis*. The calibration prior was applied with a lognormal prior, mean = 1.135 and SD = 0.8 leading to 59.0 Ma as the minimal age offset and 70.6 Ma as the 95% soft upper bound. The fossil genus †*Amyzon* (Bruner, 1991; Wilson, 1993) resolves as sister to a clade containing *Ictiobus* and *Carpiodes* (Smith, 1992). This genus was used to date the MRCA of the Ictiobinae (*Ictiobus* and *Carpiodes*) and its sister clade containing *Catostomus*, *Erimyzon*, *Hypentelium* and *Moxostoma*. A lognormal prior, mean = 0.764 and SD = 0.8 was used to set 49.4 Ma as the minimal age offset and 57.0 Ma as the 95% soft upper bound.

Two additional calibrations were also used, *Ameiurus pectinatus* (Lundberg, 1975) was used as a stem lineage calibration for the genus *Ameiurus* (*A. natalis* and *A. nebulosus*) using the include stem option in BEAST with the lognormal prior, mean=1.9 and SD=0.8 with 34.1 Ma as the minimum offset and 59.03 Ma as the 95% soft upper bound. The upper bound corresponds to the minimal age offset used in the calibration of stem Ictaluridae. This fossil was described from the Oligocene Florissant Lake Beds in Colorado, USA. It is assigned to *Ameiurus* on the basis of the broad snout and premaxillae, and the shape of the anteroventral crest of the dentary which is prominent and extends to the symphysis. It is considered to lie near the base of *Ameiurus* because the proximal posterior dentations of the pectoral spine arise from the posterior groove which is found in living *Ictalurus*. Other species of *Ameiurus* have these proximal dentations attached to the dorsal half of the spine shaft (Lundberg, 1975).

The fossil kneriid, †*Mahengichthys singidaensis* (Davis et al., 2013) was also used as a calibration. This fossil was collected from the Mahenge deposits in Tanzania which based on recovered fish fossils were assigned a Paleogene (possibly Oligocene) age (Greenwood and Patterson, 1967). A zircon crystal hypothesized to be from the eruption that created the lake has been dated using a $^{206}Pb/^{238}U$ age of 45.83± 0.17 Ma (Harrison et al., 2001) leading to estimates of the age of the fossils at 45-46 Ma. The fossil †*Mahengichthys singidaensis* is resolved as sister to the genus *Kneria* within the tribe Kneriini using a morphological matrix and also using a combined morphological and mitogenome matrix (Davis et al., 2013). Synapomorphies that support the placement of this fossil within the tribe Kneriini (extant genera *Kneria* and *Parakneria*) include the shape of the opercular bones in lateral view (squarish or square), the first six anterior epicentral bones

being highly modified and larger than the posterior ones, and the lateral line not piercing the supracleithrum. This calibration is applied with the lognormal prior mean = 2.1 and SD = 1.22 with an offset of 46 Ma and 109.1 Ma as the 95% soft upper bound.

In addition to the dating constraints, topological constraints were applied to this phylogeny in order to aid convergence. The Ostariophysians, Gonorynchiformes, Otophysi, Gymnotiformes, Cypriniformes, Characiformes and Siluriformes were each constrained to be monophyletic. The root of the phylogeny was constrained with a normal prior, mean = 245.5 and SD = 10.8, a wide prior that reflects the clade age in a phylogeny of teleost fishes (Near et al., 2012).

This analysis used the nuclear molecular markers RAG1 exon 3, ENC1 and Plagl2 with partitionfinder v1.1 (Lanfear et al., 2012) used to determine the partitioning scheme and nucleotide models using a greedy algorithm and Bayesian Information Criteria. A Bayesian phylogeny was constructed with these constraints in BEAST 1.7.5 (Drummond et al., 2012) using a birth-death prior, the analysis was repeated three times with each analysis consisting of 175,000,000 generations. Preliminary analyses suggested that a relaxed clock for each gene over-parameterized the analysis so each subset identified for the nucleotide model was assigned a separate relaxed lognormal clock. Burn in was determined in TRACER with a value of 10% (Rambaut and Drummond, 2009) and ESS values were all over 200. A maximum clade credibility (MCC) tree was calculated using Tree Annotator.

The 'Big Africa' only phylogeny was constructed using 69 taxa adding additional taxa to those from this clade in the Ostariophysian tree (Appendix 3, Table 2) and additional molecular markers were used to provide greater resolution, the single copy nuclear markers RAG1 exon 3, ENC1 and Plagl2, RAG2 and the first and second codon positions of the mitochondrial sequences CO1 and *Cytb*. This phylogeny was calibrated using the posterior probability for the 'Big Africa' clade (as defined by Sullivan et al., (2006) with the addition of *Lacuntunia enigmatica* (Lundberg et al., 2007)) from the Ostariophysian phylogeny as the prior on the root of this phylogeny. As above, partitionfinder v1.1 was used to determine the partitioning scheme using a greedy algorithm and Bayesian Information Criteria. BEAST 1.7.5 was used to generate the phylogeny using a birth-death prior, this analysis was repeated three times and run for 100,000,000 generations with each marker having a separate relaxed lognormal clock (the

mitochondrial markers had the same clock). Burn in was determined in TRACER with a value of 10% and ESS values were all over 200. A MCC tree was calculated using Tree Annotator.

4.3.3 Morphological Analyses

Nine morphological measurements, in addition to standard length, were collected from specimens in each radiation. These measurements capture shape changes from throughout the length of the specimens, and include measurements directly linked to feeding such as those that quantify the position of the mouth and eye. The morphological measurements used in this study have previously been used in studies of diverse fish systems including studies investigating morphological divergence in adaptive radiations (e.g., Alexandrou et al., 2011; López-Fernández et al., 2013; Montaña and Winemiller, 2013). 1) head length from the upper lip to the posterior edge of the operculum, 2) the eye diameter, 3) head height through the centre of the eye, 4) eye position measured as the distance between the base of the head to the centre of the eye, 5) snout length from the upper lip to the centre of the eye, 6) gape width, 7) maximum body depth, 8) depth of the caudal peduncle, and 9) body width at pectoral fin inserts. The claroteine dataset consisted of 135 specimens in 13 species, including three separate populations of *Lophiobagrus cyclurus* from Burundi, Kigoma and Zambia (Chapter 3) making 15 sample groups (minimum 1, maximum 17, median 10 samples per group). The *Synodontis* dataset consisted of 63 specimens in 8 species (minimum 2, maximum 14, median 8.5 samples per species). A full list of the specimens included in this study is given in Appendix 3 (Table 1). In order to reduce intraspecific variation, the natural logarithm of each measurement was taken, and subsequently averaged within species. These trait species averages were corrected for size while taking into account phylogenetic history using phylogenetic size correction (phyl.resid command) in the phytools R package (Revell, 2009). These residuals were then used as the input for a phylogenetic PCA in the same R package (phyl.pca command). The phylogenetic component of these commands takes into account the non-independence of trait values. Phylogenetic PCA plots can be complex to interpret as the axes show the non-phylogenetic component of shape variation but the location of the points in this space still has phylogenetic covariance (Polly et al., 2013). Due to this and to investigate intraspecific variation each analysis was

repeated without averaging within species and without the phylogenetic component in PAST (Hammer et al., 2001). Prior to the PCA size was removed as a component using the allometric vs standard option. This analysis included species not present in the molecular phylogeny leading to a dataset of 144 specimens in 16 species, again including three separate populations of *Lophiobagrus cyclurus* from Burundi, Kigoma and Zambia (Chapter 3) making 18 sample groups (minimum 1, maximum 17, median 9.5 samples per group) for the claroteine radiation, and 97 specimens in 10 species (minimum 1, maximum 14, median 12 samples per species) for *Synodontis*.

Disparity through time (DTT) plots were used to visualise morphological diversification through time (Harmon et al., 2003) using the R package GEIGER (ddt.full command) (Harmon et al., 2008). This approach calculates disparity as average pairwise Euclidean distances between species and relative disparity for each subclade as disparity in a subclade divided by total disparity. At each node the mean relative disparity was plotted, calculated as the mean of the relative disparities of all subclades that had ancestral lineages present at that point in time (Harmon et al., 2003). This analysis was repeated using the PC scores of each species for all PC axes that were considered together and for each PC axis separately. This was compared with 10,000 simulations across the MCC tree and repeated across a randomly chosen subset of 1000 trees taken from the BEAST analysis until a null hypothesis of Brownian motion with 95% confidence levels was computed (Slater et al., 2010). Following this the morphological disparity index (MDI) was calculated. The MDI is defined as the area between the line connecting disparity points from the data and the line of median disparity plots of the null model simulations, and was used to investigate if there is lower subclade disparity than expected under Brownian motion (MDI <0) or higher subclade diversity than that expected under the null model (MDI>0). In an adaptive radiation driven by ecology, disparity is expected to be distributed between subclades rather than within them. In order to take into account phylogenetic and branching time uncertainty MDI was calculated across a sample of 1000 chronograms in GEIGER with 10,000 simulations for each chronogram, and the p-value for MDI was calculated as in Slater et al., (2010). The fit of each PC axis to several likelihood models of continuous character evolution, Brownian motion,

Ornstein-Uhlenbeck model and Early Burst, was assessed on the MCC tree using the fitcontinuous command in the R package GEIGER.

4.3.4 Stable isotope analysis

White muscle was collected as this has been shown to be representative of the overall stable isotope values in fish (Hesslein et al., 1993) and used in a variety of stable isotope studies (reviewed in Dufour & Gerdeaux, 2001). A total of 113 claroteine specimens (minimum 1, maximum 22, median 10 specimens per species) and 78 *Synodontis* specimens (minimum 11, maximum 21, median 16 specimens per species) were euthanized immediately upon capture and preserved in 70% ethanol. Sub-samples were taken as controls at this stage and dried immediately. A variety of pelagic and littoral basal resources were collected to investigate the composition of sources of primary productivity that each species used. Tissue samples were oven-dried (55°C) and homogenised using a pestle and mortar. Invertebrate samples that may have been contaminated with inorganic carbon e.g., in their shells were acidified using 0.4% HCL (with controls taken). Samples were accurately weighed to 0.6mg for fish and invertebrates and 1.5mg for plants and algae in tin capsules. Samples were combusted in an ECS 4010 elemental analyser (Costech instruments) connected to a continuous flow Delta V Plus mass spectrometer (Thermo Scientific) at the Scottish Universities Environmental Research Centre (SUERC), Life Sciences Mass Spectrometry Facility, East Kilbride, UK in order to measure isotopic signatures of both nitrogen and carbon. Secondary standards (agarose gel, 14N enriched alanine, glycine and tryptophan) were used at the beginning and end of each run and every eight samples during the run to calibrate the run. The results are expressed in the standard delta notation (δ) showing the ratio of different isotopes relative to international standards (atmospheric nitrogen and Vienna-PeeDee Belemnite).

Preservation of tissues for stable isotope analysis in ethanol can be problematic. Several taxa have shown no effects of ethanol preservation on either nitrogen or carbon signatures (e.g., Barrow et al., 2008; Hobson et al., 1997) whereas other taxa have shown effects in either carbon or nitrogen signatures (e.g., Correa, 2012; Kelly et al., 2006; Xu et al., 2011). In freshwater fish the situation is complicated as differing biases have been observed in different studies, for example, *Castostomus occidentalis* was enriched in $\delta^{15}N$ 0.37 ± 0.17 (mean±

standard error, parts per thousand compared to a standard) (Sarakinos et al., 2002) whereas *Hemibarbus barbus* was depleted in $\delta^{15}N$ -0.65 ± 0.14 (Ogawa et al., 2001). Due to these changing effects average correction factors to correct for preservation effects have been unsuccessful (Kelly et al., 2006) and there have been studies which have called for species specific correction factors (Xu et al., 2011), however, in studies involving small species (e.g., *Lophiobagrus brevispinnis*) or a large number of species/populations this is often not feasible.

In order to investigate the role of preservation on isotopic signatures control samples (dried) were included for a range of fish species from different trophic levels including cichlid fishes, mastacembelid spiny eels (K. Brown, unpublished data) plus claroteine and *Synodontis* catfishes. Baseline taxa were not included as preservation biases have been found to differ between fish and molluscs (Correa, 2012). Ethanol preservation was found to have a significant effect on $\delta^{15}N$ (Wilcoxon paired test V=55, p=0.00195, t-test t=4.34, df=9, p-value=0.00187). Ethanol preservation relatively enriched nitrogen signals by on average $\delta^{15}N$ 0.454 (3 d.p.). The $\delta^{15}N$ values were corrected through the use of a bootstrapped linear correction model (2000 bootstraps) which has previously been used to correct for ethanol preservation effects in a single species (Kelly et al., 2006) but was used in this context to correct for preservation effects across taxa with the model intercept 0.467 (3.s.f.) error 0.298, gradient 0.867 (3 s.f.) error (0.0436). After this linear correction factor was applied there was no significant difference between the control and corrected values (Wilcox test, V=28 p=1, t-test t=0, df=9, p value=1, when this is plotted the intercept is not significantly different from 0 (p=1) and slope =1 (2 d.p.)).

Preservation effects had no significant effect on the $\delta^{13}C$ (Wilcoxon test V=44, p=0.106, T-test t=1.07, df=9, p=0.314). Lipids are depleted in $\delta^{13}C$ relative to other tissues so differences in lipid content can bias results. $\delta^{13}C$ values can be corrected for lipid content mathematically using the formula in Kiljunen et al., (2006). This formula is dependent upon the C:N ratio of a sample, this ratio was not significantly different between control samples and those stored in ethanol (Wilcoxon test V=12, p=0.25, t-test t=-1.01, df=8, p=0.343) meaning these mathematical corrections could be applied.

Nutrient regimes differ from the north to south of LT (Plisnier et al., 1999). In order to compare between sites, accurate isotopic baselines from different taxa

feeding at different points of the food web are required, in order to correct for these differing baseline isotopic values. A lack of available samples from the northern basin meant that this was not possible for this study, and so only samples from the southern (Zambian) sites were included in the isotopic analyses. These sites were sampled most intensively at Sumbu in the west and Mpulungu in the east of LT. The isotope baselines between the two sites are similar, including those for bivalves, which are more depleted in $\delta^{13}C$ at lower taxonomic levels than the other taxa sampled (Appendix 3, Figure 4). Based on these values the samples from each location were pooled and analysed together.

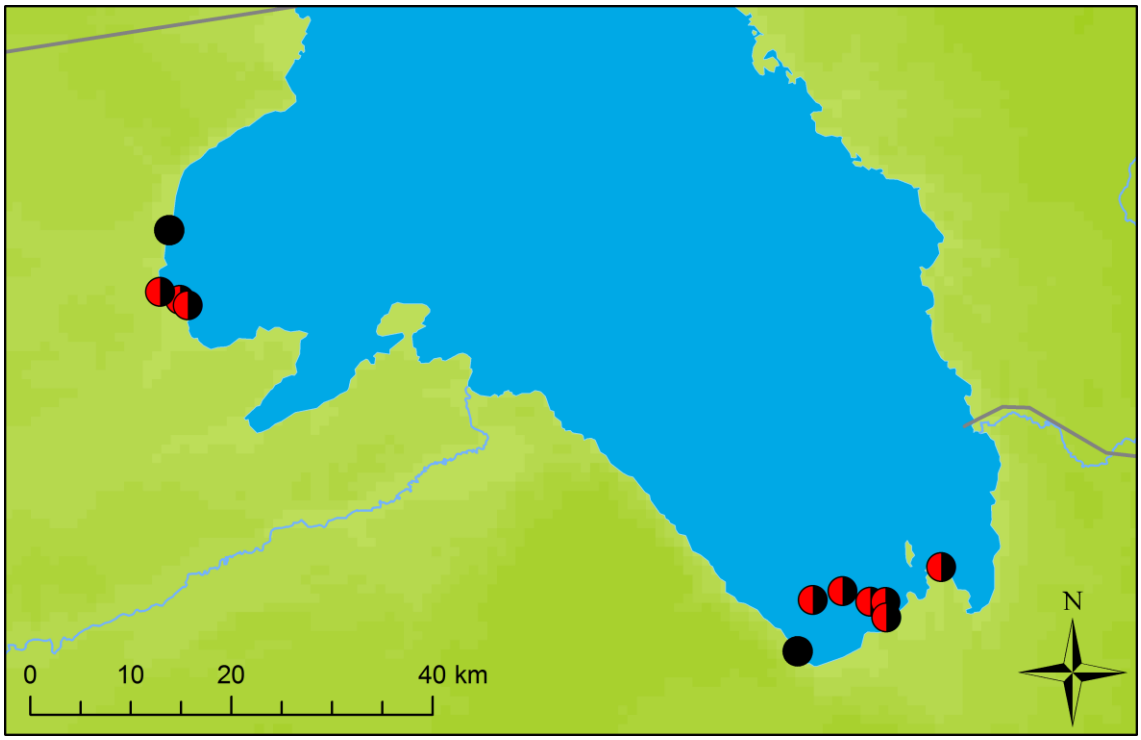


Figure 1 Sampling sites in Zambia. Sites where only claroteine samples were collected are shown in black, sites where samples from both radiations were collected are shown as half red half black.

The range of $\delta^{15}N$ was used to represent the trophic partitioning of the samples, as $\delta^{15}N$ is enriched at higher trophic levels, and $\delta^{13}C$ was used to estimate the contribution of littoral and pelagic basal resources (Post, 2002). The R package Stable Isotope Analysis in R (SIAR, Parnell et al., 2010) was used to calculate standard ellipses for each species in order to assess isotopic niche space and overlap between species.

The phylogenetic signal for $\delta^{15}N$ and $\delta^{13}C$ signatures on the MCC tree was assessed for the claroteine radiation using Blomberg's K (Blomberg et al., 2003), implemented using the multiPhylosignal command in the R package picante (Kembel et al., 2010) with p-values calculated based on observed vs random

variance in phylogenetic contrasts. The low sample numbers in *Synodontis* make testing for phylogenetic signal difficult using the isotope data. Disparity through time plots were used as described above for the PC axes to investigate patterns in isotopic niche through time and also signatures of $\delta^{15}N$ and $\delta^{13}C$ separately. These plots were produced for the *Synodontis* radiation in order to provide a comparison to the morphology plots, however, the low number of species sampled mean that these plots must be interpreted with caution. The relationship between morphology and ecology was assessed using phylogenetic generalized least squares regression using the R packages GEIGER to simulate Brownian Motion correlation structure (Harmon et al., 2008) and nlme (Pinheiro et al., 2014) (gls command) with mean $\delta^{15}N$ against retained PC axes 1-4 from the phylogenetic PCA and separately $\delta^{13}C$ signatures against retained PC axes 1-4 from the phylogenetic PCA over 1000 trees pruned to include only taxa for which morphological and isotope data were available. The analysis was performed for the claroteine radiation but not the *Synodontis* radiation due to the low number of species for which isotope data were available. The correlation between claroteine isotope signatures and morphology was further investigated using phylogenetic canonical correlation analysis (command phyl.cca in the phytools R package) to investigate correlations between PC axes 1-4 from the phylogenetic PCA and isotope signatures (both $\delta^{15}N$ and $\delta^{13}C$ signatures). This method finds the axes of largest correlation and assesses the significance of this correlation.

4.4 Results

4.4.1 Topology and dating of the phylogeny

The model and partitioning scheme for the ostariophyisan phylogeny was a partitioning scheme with two subsets, the first containing the first and second codon positions for all genes with the model GTR+I+G and the second subset containing the third codon partitions for all genes with the model SYM+I+G. In this phylogeny the Siluriformes had an age of 143.58 Ma (95% HPD 120.91-163.09) which is older than estimates from a fossil calibrated phylogeny of all ray-finned fishes based on sequences from nine nuclear genes (106.1 95% HPD 89.9-123) (Near et al., 2012) but younger than an estimate based on a fossil calibrated phylogeny built from mitogenomes (180Ma, 95% HPD 162-198) (Nakatani et al.,

2011). The ostariophysian tree (Appendix 3, Figure 3) resolved the 'Big Africa' clade (Sullivan et al., 2006) with the addition of *Lacuntunia enigmatica* (Lundberg et al., 2007) with an age estimate of 83.83 Ma (95% HPD 66.76-93.16). This age estimate translated into a normal prior on the root of the 'Big Africa' only phylogeny with a mean of 80 Ma and a standard deviation of 8.0 Ma, truncated to only include 55-150 Ma. In the 'Big Africa' only analysis a partitioning scheme was chosen with two subsets, one with the first and second codon positions for all genes (including the mitochondrial genes) with the model SYM+I+G and the second subset containing the third codon position for all genes with the model SYM+G. In this analysis (Figure 2) the claroteine LT radiation is dated at 5.78 MA (95% HPD 3.87-7.97) and the *Synodontis* LT radiation (including *S. victoriae* which is found within this clade but not in LT) at 2.7 Ma (95% HPD 1.67-3.94). The broad relationships within the LT claroteine radiation are the same as those shown in the combined nuclear and mitochondrial analysis in Chapter 2 (Peart et al., 2014) with a few minor differences. In Chapter 2 *Bathybagrus tetranema* is sister to the LT *Chrysichthys* taxa, however in this analysis it is sister to a clade containing both *Phyllonemus* and LT *Chrysichthys*. Relationships between the three species of LT *Chrysichthys* (with the exception of *C. sianenna* that does not differ in its placement) also differ between analyses and in this analysis *Lophiobagrus cyclurus* is paraphyletic with *L. aquilus* within this clade. There are also differences in the topology of the LT *Synodontis* clade (between Clades 1 - 6) from this analysis and previous analysis using only mitochondrial data (Day and Wilkinson, 2006; Koblmüller et al., 2006; Appendix 3, Figure 1) and those including nuclear sequences (Day et al., 2009). This phylogenetic uncertainty is reflected in low support values between species (Figure 2). The inclusion of a subset of 1000 trees, where possible, in disparity analyses is used to account for this phylogenetic uncertainty.

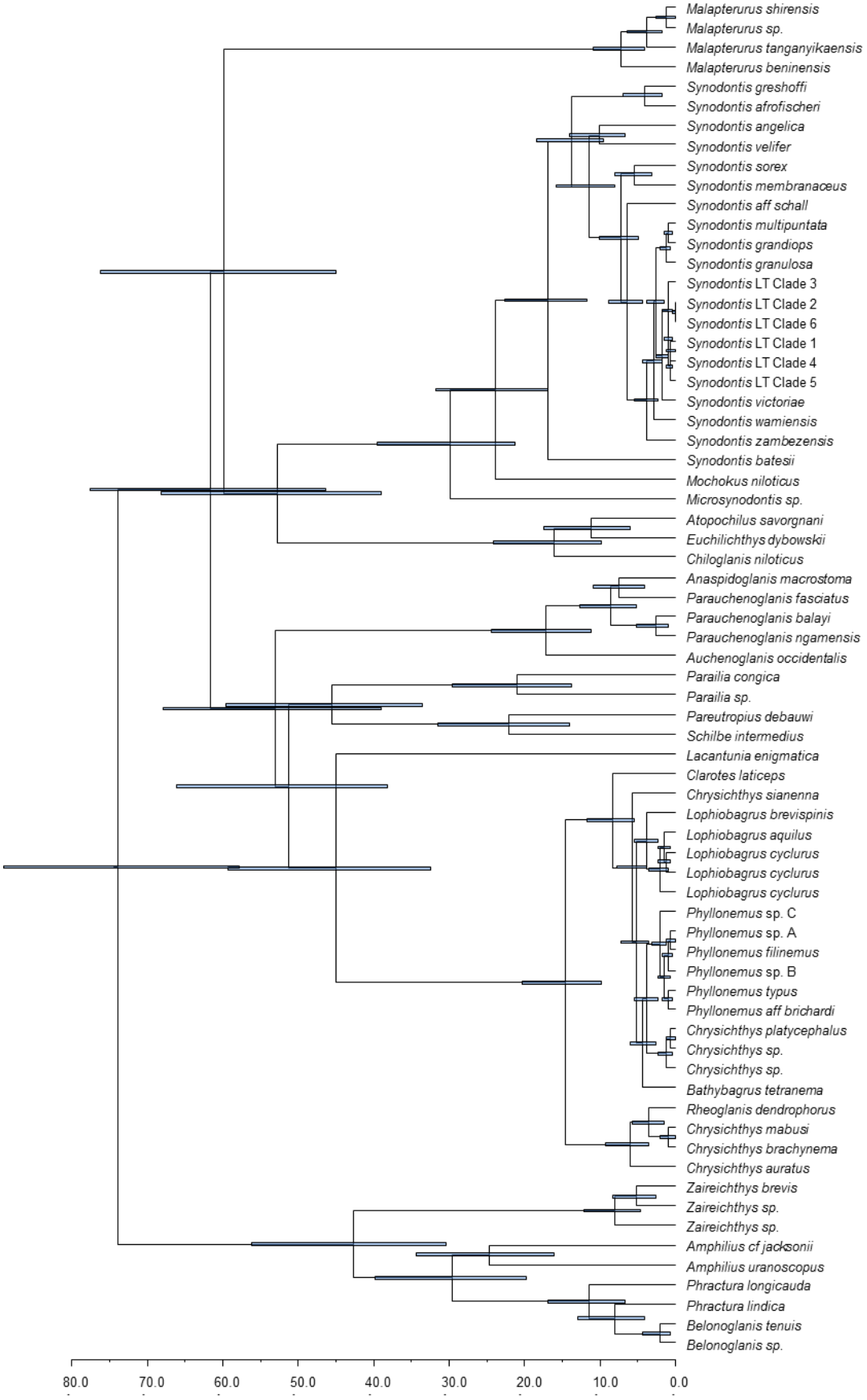


Figure 2 BEAST analysis for the 'Big Africa' tree using both nuclear and mitochondrial markers. Scale bar is Ma. Node bars represent 95% confidence intervals around the node ages.

4.4.2 Morphology

In the claroteine radiation the first four phylogenetic PC axes explained over 90% of the variation with 41.5%, 27.8%, 14.2% and 8.49% respectively. Species with negative values on PC1 show greater relative head heights, higher eye heights, larger gape width and greater body depth, caudal peduncle depth and body width. Species with a high score on PC2 have a longer head length and smaller eye diameter while high scores on PC3 mainly reflect a long snout length. In the claroteine radiation the different genera are separated in morphospace by PC1 and to a lesser extent by PC3 (Figure 3). Within the taxa whose phylogenetic position was unclear in Chapter 2 (Peart et al., 2014) *C. sianenna* and *B. tetranema*, *C. sianenna* does not cluster with other taxa in morphospace whereas *B. tetranema* occupies a similar area of morphospace to the LT *Chrysichthys* taxa along the first three PC axes. The *Lophiobagrus* taxa are less clustered along PC2 and PC3 with *L. aquilus* occupying a similar area of morphospace to the *Phyllonemus* taxa (Figure 3). The two small species (*L. brevispinis,* 40.66mm max standard length in this study and *Phyllonemus* sp. C, 46.36mm max standard length) do not occupy the same areas of morphospace based on these analyses but are similar to the other members of their respective genera on PC1.

In the *Synodontis* dataset the first three PC axes explain over 97% of the variation with 77.4%, 14.7% and 5.2% respectively. The high scores on the first PC axis correspond to smaller head lengths and head height with a lower eye position, smaller eyes, shorter snout length and smaller body depth. On the second PC axis high scores correspond to larger caudal peduncles with a lesser relationship to larger body widths. High scores on the third PC axis represent mainly larger gape width and larger eye diameter. The *Synodontis* taxa are spread evenly across PC1 whereas PC2 and PC3 show some distinct groups with *S. multipunctatus* distinct from other taxa especially on PC2. Clade 3 and *S. granulosus* show very similar scores on PC2 with the remaining taxa tightly clustered on both PC2 and PC3.

In the non-phylogenetically corrected claroteine dataset the loadings are similar to the phylogenetically corrected analysis for PC1 (though inverted) whereas those for PC2 correspond to the loadings on PC3 for the phylogenetically corrected analysis and vice versa. In the non-phylogenetically corrected claroteine dataset the greatest intraspecific variation is in *Lophiobagrus asperispinis* (Figure 4) for which genetic data is not available so it is not included in the phylogeny. Only four

specimens were available for this species including the holotype and paratopotype with all four widely distributed across PC1 and PC2, indicating considerable variation among specimens assigned this name. In general intraspecific variation is small compared to interspecific variation. Unlike in the phylogenetically corrected analyses (Figure 3) *C. sianenna* is not separated in morphospace on PC1 or PC2 (Figure 4) with only a slight separation on PC3 (Appendix 3, Figure 4). There are some general patterns with the genera clustering together on PC1 and PC2 though *L. brevispinis* is separate from the remaining *Lophiobagrus* species and there is some separation between *Phyllonemus* sp. C and the other *Phyllonemus* species. *Bathybagrus tetranema* clusters with the LT *Chrysichthys* species as it does in the phylogenetically corrected analysis (Figure 3).

In the non-phylogenetically corrected *Synodontis* dataset, loadings of each morphological measurement in the principal component analysis are similar to those in the phylogenetically corrected analysis. There is more overlap between taxa on the PCA plots (Figure 4) in the *Synodontis* dataset than seen in the claroteines. *Synodontis multipunctatus* shows the least intraspecific variation on the first two PC axes and only overlaps in morphospace with the two taxa to which it is proposed to be most closely related, *S. grandiops* and *S. granulosus*. *Synodontis granulosus* has the largest intraspecific variation overlapping in morphospace with five different taxa shown by convex hulls (Figure 4). In this analysis Clade 2 and Clade 5 overlap, as do Clade 6 and Clade 3 on PC axes 1 and 2, which does not reflect the relationships seen in the non-phylogenetically corrected PC axis (Figure 3). Similar patterns are seen when PC1 and PC3 are plotted together with *S. multipunctatus* and *S. grandiops* overlapping only with *S. granulosus* which also overlaps with Clade 2 and Clade 3 (Appendix 3, Figure 5). Each of the unnamed clades overlap in morphospace.

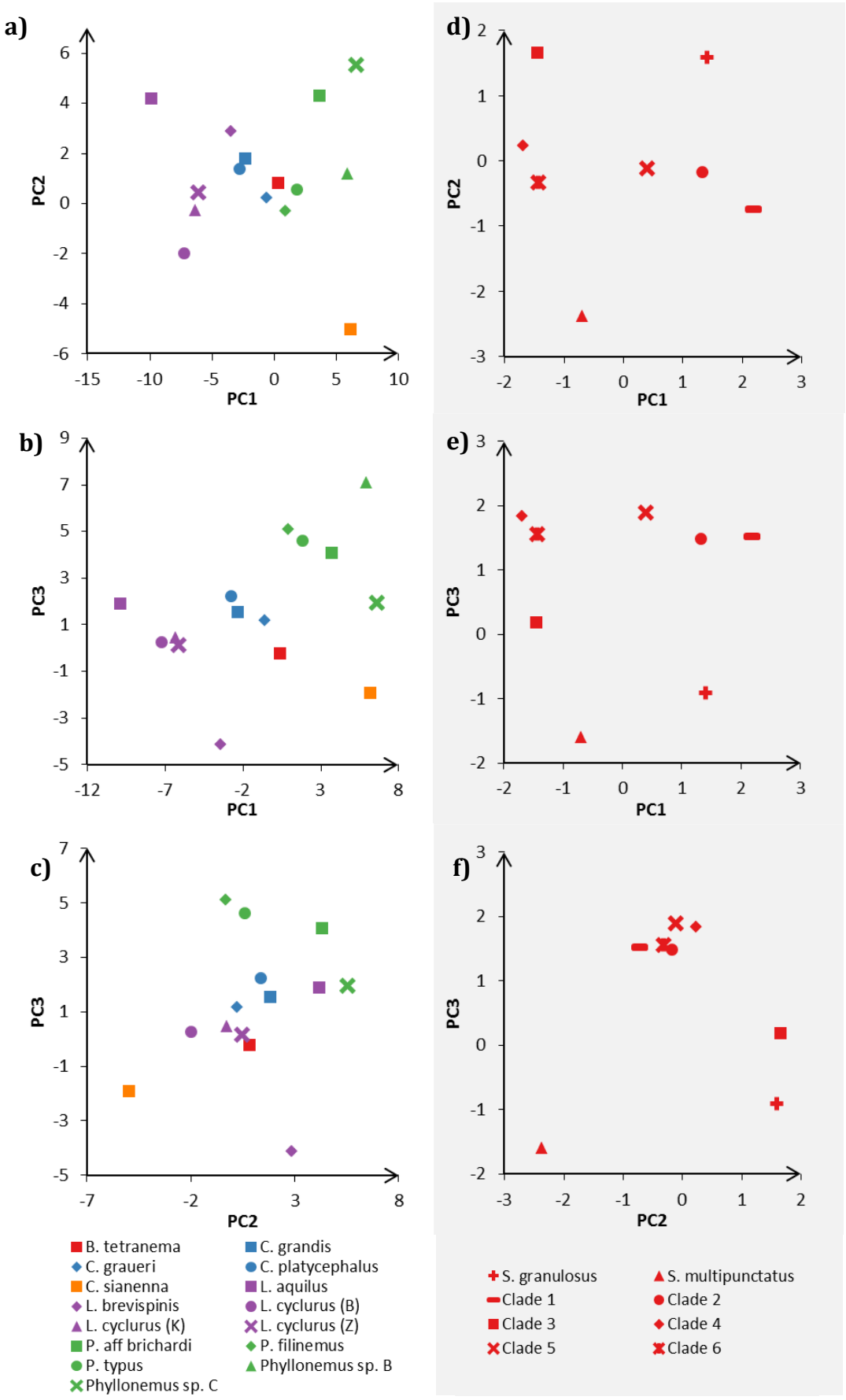


Figure 3 Phylogenetically corrected PCA plots showing the first three axes for the claroteine (a, b, c, white background) and *Synodontis* (d, e, f, grey background) with legends shown below the plots.

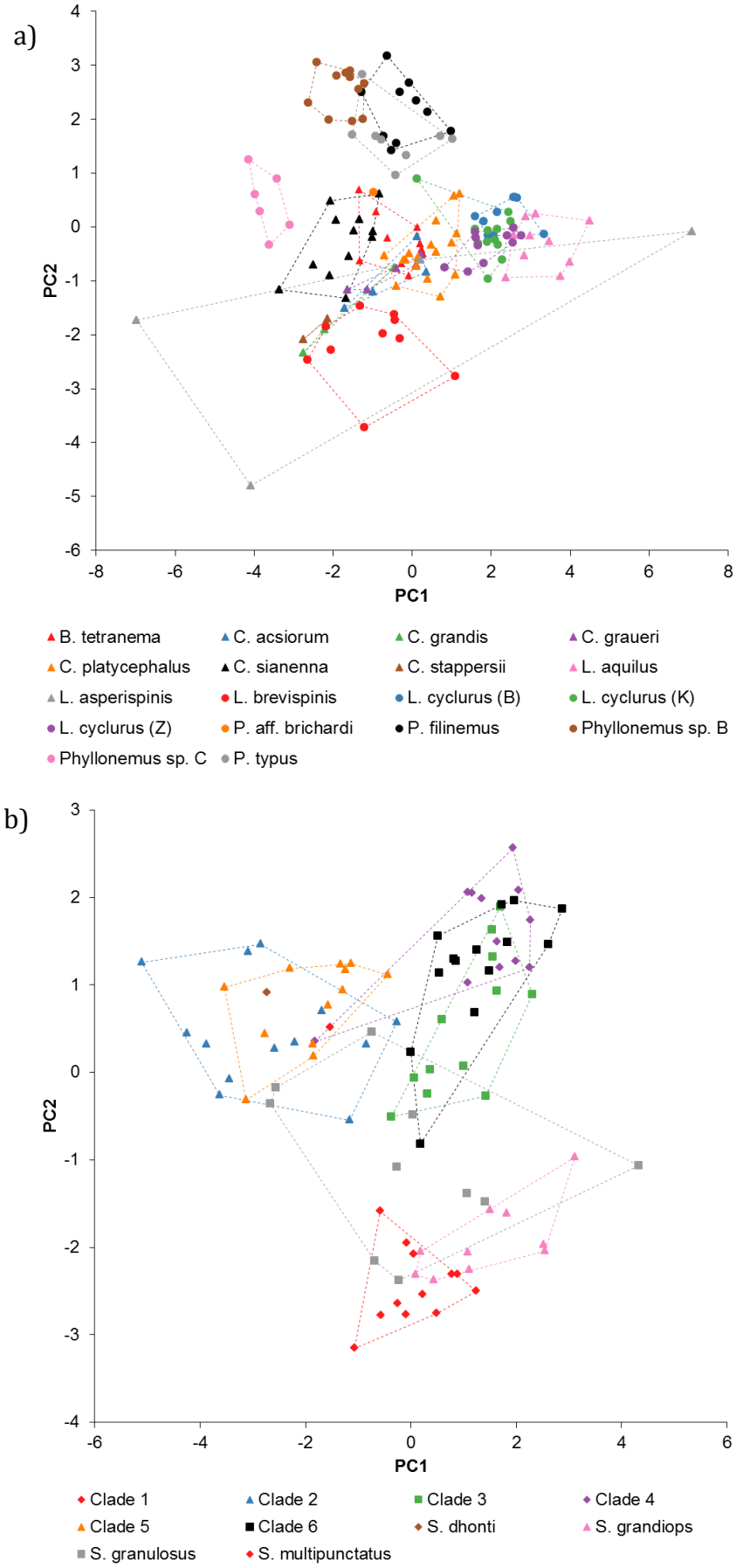


Figure 4 Non-phylogenetically corrected PCA plots for PC axes 1 and 2 for all individuals in the claroteine (a) and *Synodontis* (b) datasets. Convex hulls are drawn for taxa with more than 2 specimens.

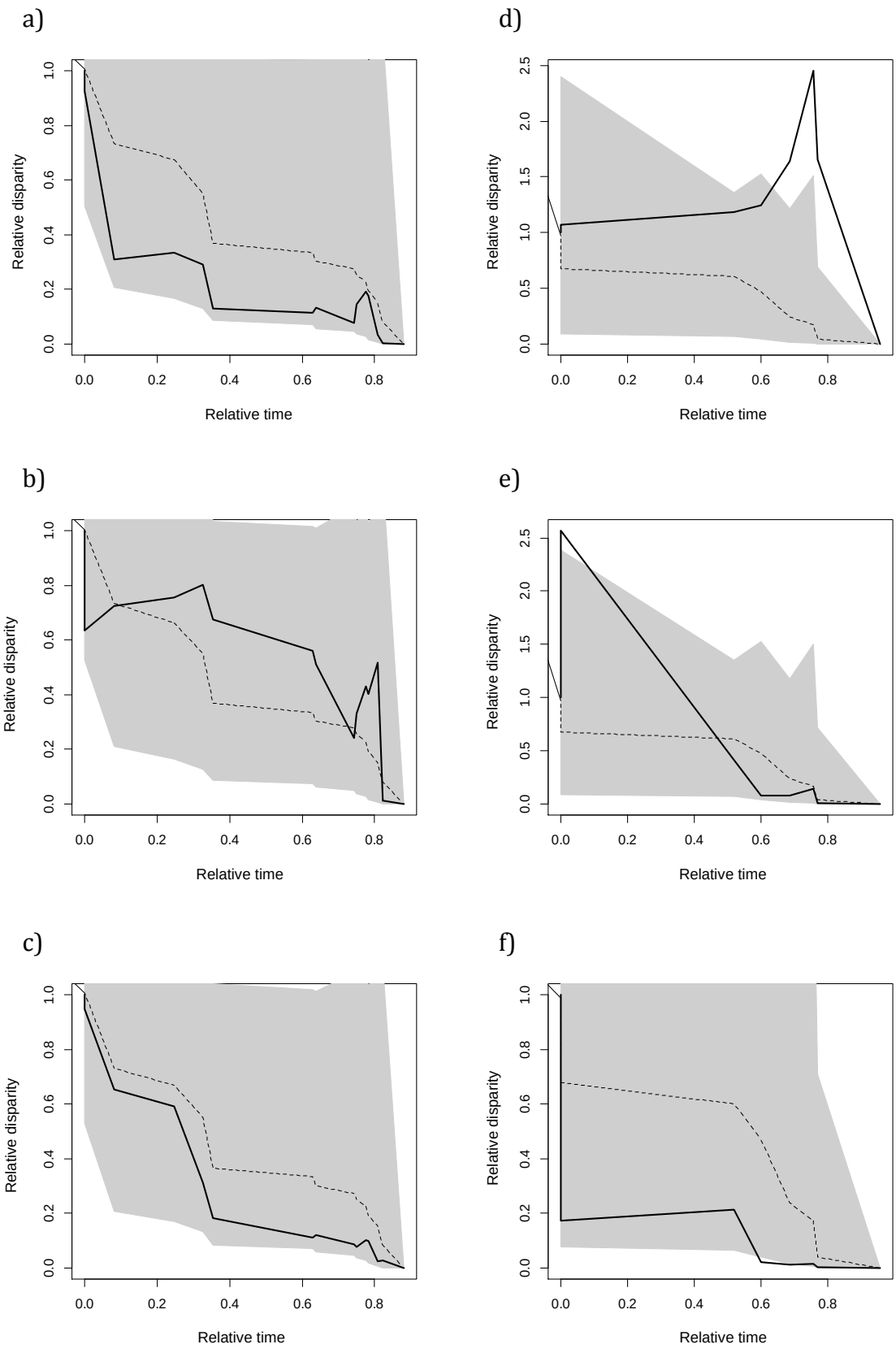


Figure 5 Disparity through time plots for the claroteine (a – PC1, b – PC2, c – PC3) and *Synodontis* (d – PC1, e – PC2, f – PC3), based on the MCC tree, grey area shows 95% confidence intervals, black line-disparity, dashed line predicted disparity under Brownian motion. The relative time axis shows the proportion of time from the origin of the radiation (0) to the present (1).

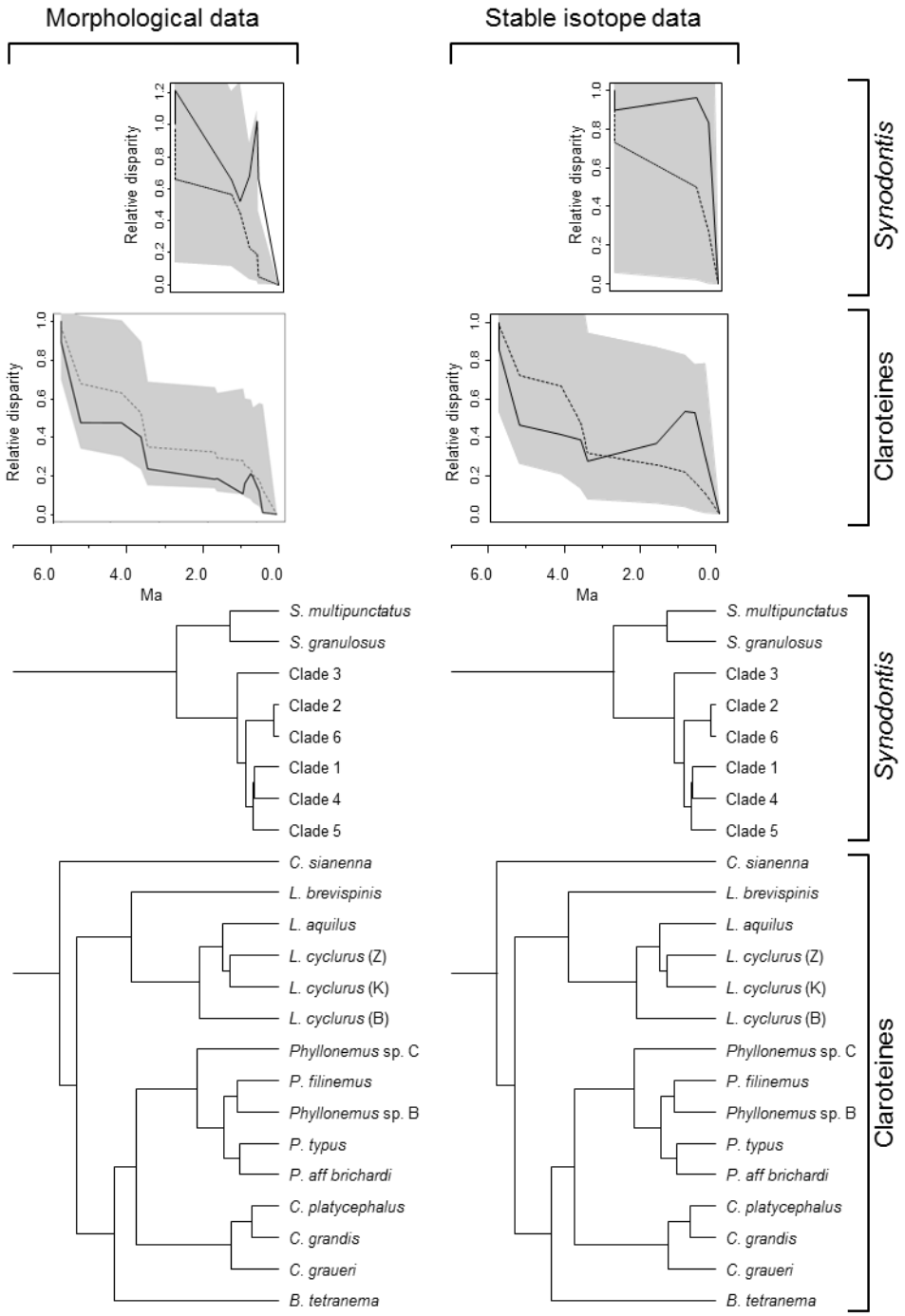


Figure 6 Morphological and stable isotope disparity through time (DTT) plots, and dated phylogenetic trees for *Synodontis* and the claroteines displayed on the same time scale. The morphological DTT plot for *Synodontis* shows phylogenetically corrected PC axes 1-3. The morphological DTT plot for the claroteines shows phylogenetically corrected PC axes 1-4. The most recent common ancestor of the claroteines and *Synodontis* is older than the scale shown.

In the claroteine disparity through time (DTT) plots when the first four PC axes are considered together the result is similar to Brownian motion (Figure 6) but there is a peak in subclade relative disparity towards the present which is strongest in PC2 (Figure 5b) which generally shows higher relative disparity than that expected under Brownian motion. This peak towards the present occurs at the same time as a much larger peak subclade relative disparity seen in the *Synodontis* radiation (Figure 65; Figure 6), which is mostly influenced by PC1. This is reflected in the MDI value (Table 1), which is significantly larger than that seen under Brownian motion on PC1 for the *Synodontis* radiation ($p<0.001$ MCC tree, $p= 0.02 \pm 0.028$ SD across 1000 trees). This is in contrast to a MDI lower than that expected under Brownian motion for PC1 in the claroteine radiation, though this is not significant ($p= 0.063$ MCC tree, $p= 0.066 \pm 0.028$ SD across 1000 trees), which is also seen in PC3 of the *Synodontis* dataset (Figure 5f; Table 1). Akaike weights suggest a model of Brownian motion for each PC axis on the MCC tree in the claroteines, whereas in the *Synodontis* radiation Akaike weights suggests an OU model for PC1 (Table 2) and Brownian motion for PC2 and PC3.

Table 1 MDI for each PC axes using the MCC tree and 1000 trees

a) Claroteinae

	MCC tree		1000 Trees			
	MDI	**P-value**	**MDI mean**	**MDI standard deviation**	**P-value mean**	**P-value standard deviation**
PC1	-0.209	0.063	-0.197	0.022	0.066	0.028
PC2	0.106	0.305	0.156	0.062	0.246	0.084
PC3	-0.125	0.190	-0.112	0.030	0.214	0.068
PC4	-0.035	0.391	0.006	0.070	0.386	0.090
PC1-4	-0.114	0.094	-0.099	0.028	0.132	0.073

b) Synodontis

	MCC tree		1000 Trees			
	MDI	**P-value**	**MDI mean**	**MDI standard deviation**	**P-value mean**	**P-value standard deviation**
PC1	0.706	0.000	0.738	0.111	0.020	0.029
PC2	0.381	0.123	0.610	0.326	0.100	0.109
PC3	-0.314	0.062	-0.256	0.070	0.109	0.085
PC1-3	0.307	0.118	0.396	0.113	0.093	0.057

Table 2 Akaike weights for each PC axes

a) Claroteinae

	Brownian motion		Ornstein-Uhlenbeck		Early burst	
	dAICc	Akaike Weight	dAICc	Akaike Weight	dAICc	Akaike Weight
PC1	0.000	0.693	3.309	0.133	2.764	0.174
PC2	0.000	0.610	1.609	0.273	3.309	0.117
PC3	0.000	0.705	3.309	0.135	2.959	0.161
PC4	0.000	0.723	3.309	0.138	3.302	0.139

b) Synodontis

	Brownian motion		Ornstein-Uhlenbeck		Early burst	
	dAICc	Akaike Weight	dAICc	Akaike Weight	dAICc	Akaike Weight
PC1	3.577	0.143	0.000	0.853	10.577	0.004
PC2	0.000	0.934	6.409	0.038	7.000	0.028
PC3	0.000	0.901	7.000	0.027	5.071	0.071

4.4.3 Trophic partitioning

Standard areas of the ellipses were larger for the *Synodontis* radiation than the claroteine radiation (Table 3; Figure 7) though *C. platycephalus* for which a large number of samples were collected also had a large ellipse area. In particular, there was greater spread in terms of carbon sources in the *Synodontis* radiation (Figure 7; Figure 8), which exploit more littoral resources, compared to the more pelagic resources mostly exploited by the claroteine.

Table 3 Standard area of ellipses using both carbon and nitrogen data.

Taxon	Standard area of ellipse
B. tetranema	0.09
C. platycephalus	4.12
C. sianenna	0.28
L. aquilus	2.75
L. brevispinis	2.62
L. cyclurus (Z)	0.57
P. aff. brichardi	1.26
Phyllonemus sp. C	1.23
P. typus	1.68
Clade 3	1.70
Clade 4	10.30
Clade 5	4.58
Clade 6	3.95
S. multipunctatus	4.98

In the claroteine radiation there is a high level of niche overlap between *L. aquilus* and *L. brevispinis* despite the difference in size between these species (max size *L. aquilus* 67.04 mm compared to max size *L. brevispinis* 40.66mm in this study) but niche overlap is far higher in the *Synodontis* radiation (Figure 7; Figure 8; Table 4). There is evidence of niche partitioning within the genus *Phyllonemus* with *P. typus* not overlapping with either of the *Phyllonemus* species with which it was collected in sympatry (*Phyllonemus* sp. C and *P. aff. brichardi*).

Table 4 Overlap of standard ellipse values.

a) Claroteinae

	B. tetranema	C. platycephalus	C. sianenna	L. aquilus	L. brevispinis	L. cyclurus	P. aff. brichardi	Phyllonemus sp. C	P. typus
B. tetranema	X	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
C. platycephalus		X	0.000	0.000	0.363	0.000	0.208	0.869	0.000
C. sianenna			X	0.000	0.000	0.000	0.000	0.000	0.000
L. aquilus				X	0.969	0.508	0.101	0.000	0.751
L. brevispinis					X	0.237	0.804	0.367	0.000
L. cyclurus						X	0.000	0.000	0.000
P. aff. brichardi							X	0.310	0.000
Phyllonemus sp. C								X	0.000
P. typus									X

b) Synodontis

	Clade 3	Clade 4	Clade 5	Clade 6	S. multipunctatus
Clade 3	X	0.439	1.455	0.000	0.000
Clade 4		X	1.235	2.851	0.570
Clade 5			X	0.065	0.267
Clade 6				X	0.000
S. multipunctatus					X

The K values of phylogenetic signal in the claroteine radiation were 0.755 for nitrogen and 0.844 for carbon with both nearing significance based on p-values calculated based on observed vs random variance in phylogenetic contrasts (p= 0.072 and p=0.083). In the claroteine disparity through time plots the pattern with the isotope data is similar to that seen in the morphological data, with an initial decrease compared to Brownian motion followed by a peak in disparity towards the present (Figure 6). This pattern is seen in both the carbon and nitrogen signatures (Appendix 3, Figure 6). These results are not significantly different from Brownian motion based on MDI values (Appendix 3, Table 3). The *Synodontis*

disparity through time plots are more difficult to interpret due to the low taxon numbers that were sampled from Zambia but these results too are not significantly different from that expected under Brownian motion (Figure 6; Appendix 3, Figure 6). The disparity through time graph for both isotopes combined show that subclade disparity is maintained at a high level until the present when this drops steeply. This pattern is mainly driven by the carbon signatures with nitrogen showing a pattern similar to that observed in the combined morphology dataset (Appendix 3, Figure 6). The claroteine phylogenetic generalized least squares regression across 1000 trees showed no significant relationships between either $\delta^{15}N$ or $\delta^{13}C$ signatures with the correlation between $\delta^{13}C$ signatures and PC2 closest to significance (mean p=0.099, SD=0.042). These results are corroborated by the phylogenetic canonical correlation analysis, which was also not significant (p=0.402, p=0.809).

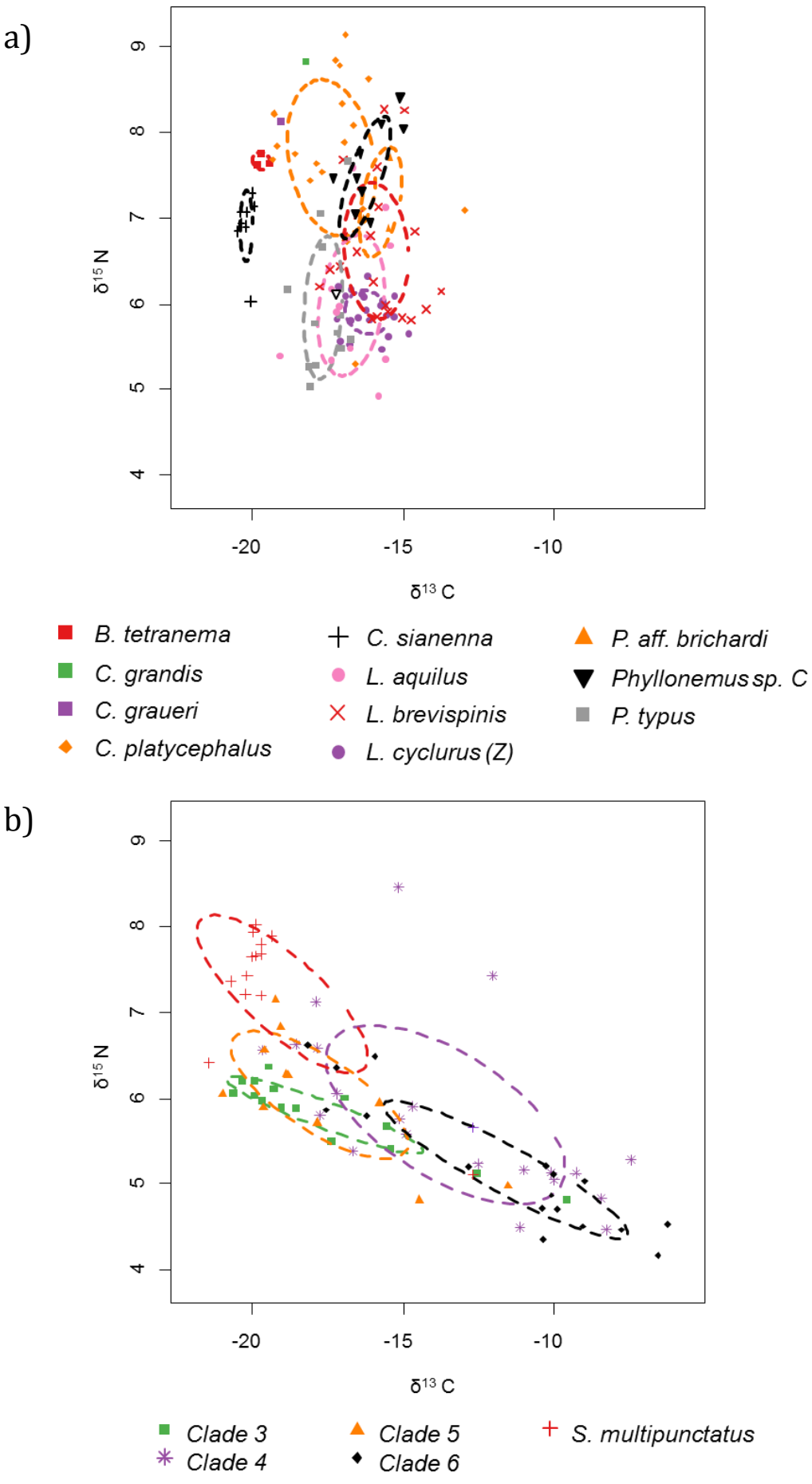


Figure 7 Scatter plots of stable isotope values for the claroteines (a) and *Synodontis* (b). Ellipses show the standard ellipse area of each species.

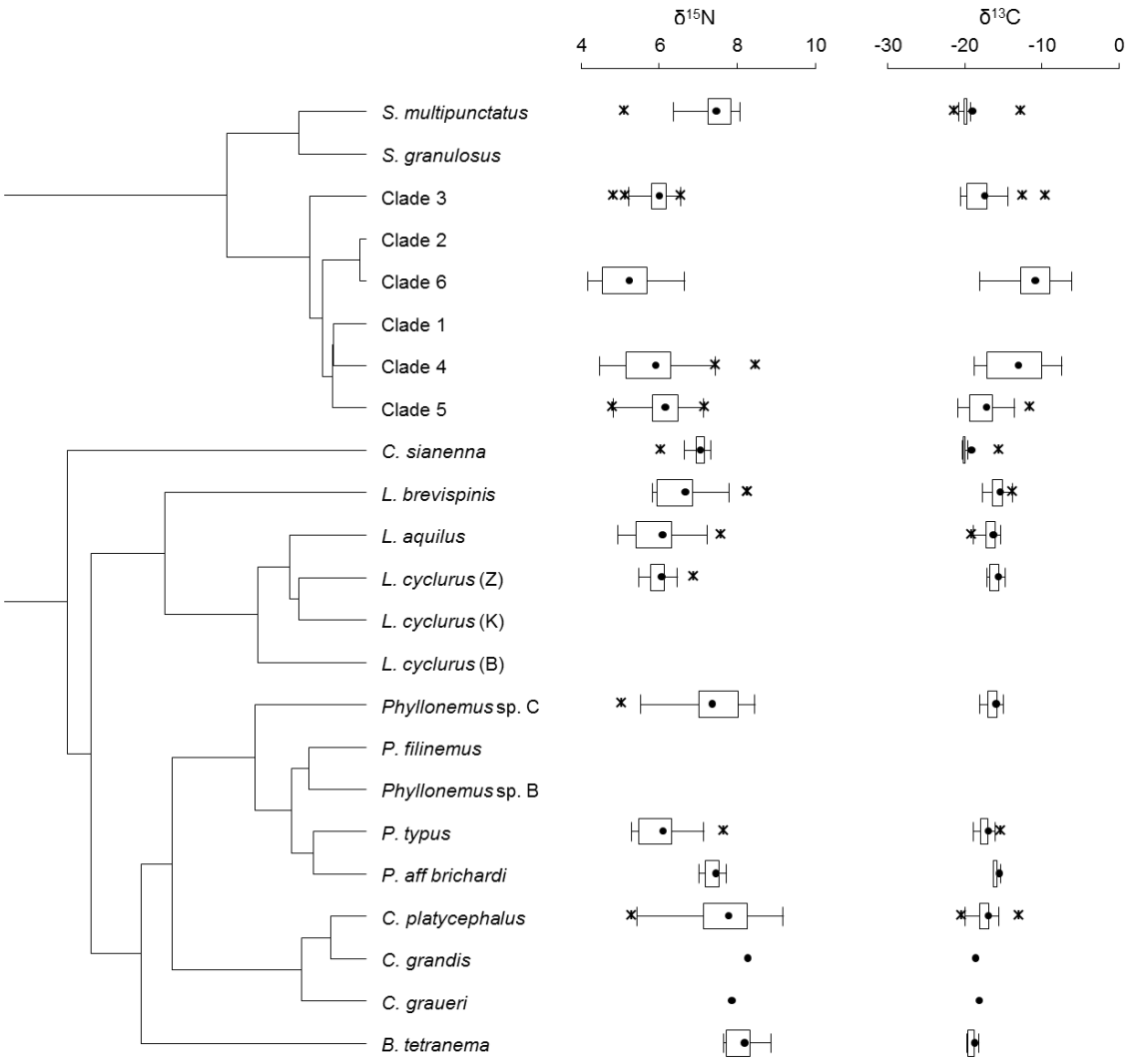


Figure 8 Phylogenetic tree of both the claroteine and *Synodontis* showing the carbon and nitrogen values for each species sampled in Zambia. Dot show the mean value and crosses show outliers greater than 1.5 x interquartile range.

4.5 Discussion

This study uses a robustly calibrated tree to place two sympatric catfish radiations in the same phylogenetic context and contrast the different adaptive patterns seen. The *Synodontis* display patterns of diversification more similar to each of the genera within the claroteine radiation than this radiation as a whole. In the claroteine radiation there is evidence of adaptive divergence between genera in the disparity through time plots and evidence of both trophic partitioning and niche overlap within genera. In *Synodontis,* species have larger isotopic niches and there is evidence of morphological conservatism with greater similarities between subclades than expected under the null model.

4.5.1 Diversification of species in each radiation

There are similarities between the two LT catfish radiations, with geographically restricted species found in each radiation. In the Claroteinae, geographic restriction is seen in *Lophiobagrus* and *Phyllonemus* (Chapter 2; Chapter 3) and possibly in LT *Chrysichthys* (discussed in Hardman, 2008). The extent of geographic restriction in *Synodontis* is unclear (discussed in Wright and Page, 2006) but there is some evidence in this dataset which included sampling from Burundi, Tanzania and Zambia, for example, Clade 2 is only sampled from the northern basin whereas Clade 3 and Clade 4 are only sampled from the southern basin.

The confidence intervals around the node ages in the phylogeny, with many branching events close together, make it difficult to compare the synchronicity of diversification events between the radiations. However, placing both radiations in the same phylogenetic context does allow timing events in each to be compared (Figure 2) and, despite limitations, does provide an improvement on the problems of comparing two small radiations both dated using different single calibrations discussed in Chapter 2 (Peart et al., 2014). In this analysis the onset of diversification in *Synodontis* is much younger, at 2.7 Ma (1.67-3.94 Ma), than previous dates for the age of the radiation calculated using single fossil calibrations. For example, the radiation was dated at 7.9Ma (5.7-10 Ma) in a study including sampling from throughout the genus *Synodontis* with a single calibration (Day et al., 2013). In contrast, this study estimates the onset of diversification in the claroteines at 5.78 Ma (3.87-7.97 Ma) which is similar to the age estimate of

5.08 (3.61-6.84 Ma) in Chapter 2 (Peart et al., 2014) using a different fossil calibration, only nuclear markers (including the non-coding S7 intron) and additional taxa from *Chrysichthys*. These results suggest that the catfishes of LT were diversifying at similar times, the origin of the *Synodontis* radiation coincides with the origin of the LT claroteine genera and diversification events in the *Synodontis* clades are contemporaneous with diversification in LT *Chrysichthys*, *Phyllonemus* and *Lophiobagrus*.

The younger age estimate of the LT *Synodontis* radiation could be related to the small amount of *Synodontis* sampling outside of the LT radiation in this study, different molecular markers or the unbalanced nature of the phylogeny. In this study, each LT radiation is sampled at the species level, however the rest of the 'Big Africa' clade is represented at a higher taxonomic level, with additional sampling outside the Mochokidae and Claroteinae similar to that from the study in which the clade was named (Sullivan et al., 2006). Despite the limitations of this method for dating the radiations it provides a way of comparing radiations separated over a large stretch of evolutionary time, for which a species level phylogeny with multiple calibrations difficult is to obtain.

It is worth noting that *Lacantunia enigmatica*, a Mesoamerican catfish species that resolves within the 'Big Africa' phylogeny, is dated at 54.32 Ma (35.11-70.62) in the ostariophysian phylogeny in this study. Using the same sequence data for *L. enigmatica*, Lundberg et al., (2007) dated this species at 75-94 Ma. However, the claroteid calibration used in that analysis is problematic (as discussed in Chapter 2), therefore the different constraints used in this study perhaps provide a more robust age estimate of its divergence from its African relatives despite the wide prior used to calibrate the root of the 'Big Africa' phylogeny (reflecting the uncertainty in the ostariophysian analysis).

4.5.2 Resource partitioning

The *Synodontis* exploit a greater range of carbon inputs than the claroteine both within each species and within the radiation, and have a more littoral signature suggesting that near shore production, for example algae, is the dominant source of carbon in this genus. The *Synodontis* in general show greater niche overlap than the claroteine, with the greatest niche overlap observed between Clade 4 and Clade 6 (which are enriched in $\delta^{13}C$ signatures relative to the other taxa). Both of

these taxa possess a hind gut, though show considerable variation in intestine length, suggesting that there are morphological adaptations in this group to exploit different resources. The fact that these two taxa are not sister to each other suggests that this may be a case of convergent evolution. In the claroteine radiation there are different patterns in resource partitioning. Most species occupy a narrow range of $\delta^{13}C$ values (Figure 8) and there is some niche overlap both within and between genera (Figure 7; Figure 8). However there is also some evidence of niche partitioning, for example within the genus *Phyllonemus*.

Stable isotope signatures of white muscle are incorporated over an extended period of time (Hesslein et al., 1993). This is useful as gut content analysis, the most common alternative method of dietary analysis, provides only a snapshot of the last meal and requires many specimens, especially as some will have empty stomachs. However, investigating stable isotope signatures as a proxy for trophic niche has limitations, the exact prey cannot be determined only their approximate place in the food web, and small differences between species cannot be distinguished. In addition, overlap of isotopic niches does not imply direct competition between species, for example, in LT shrimp eating cichlids species eat overlapping size classes of shrimps (which would be predicted to cause identical isotopic signatures) but partitioning occurs via different capturing techniques (Yuma et al., 1998). This is relevant in this study as the claroteine and *Synodontis* are thought to forage at different times of day, with the claroteine nocturnal and *Synodontis* diurnal, and therefore not to be in direct competition. In addition, species may forage in different habitats, for example, LT *Chrysichthys* species are thought to segregate by substrate type with some preferring sand whereas other are found over rocks (Coulter, 1991). Although there is some overlap in isotopic signatures between each radiation (Figure 8) the differences in $\delta^{13}C$ signatures suggest that they do not consume the same food resources. In order to properly assess the place of each species from these radiations in the food web of LT a far larger study would be required with greater taxonomic sampling of possible prey and predators and potential competitors.

The samples for the isotope analysis investigating niche partitioning were collected from sites in Zambia (Figure 1) in the southern part of LT, which has a different nutrient cycling regime to the northern basin, including wind driven upwelling (Coulter and Spigel, 1991; Naithani et al., 2003) which adds more

nutrients to the food web. Elevated nutrient levels are important in the pelagic food web (O'Reilly et al., 2002). The presence of upwelling can complicate the interpretation of stable isotope signatures in the pelagic food web as the isotopic signatures are diluted by time averaging in the upper trophic levels (O'Reilly et al., 2002). Increased sampling from additional geographic locations would be useful as competition is a powerful force through competitive exclusion leading to niche partitioning when nutrients are limited, so perhaps species which show overlap in isotopic niches in Zambia may show greater resource partitioning in other locations with lower nutrient inputs. However, the availability of food does not influence the amount of niche overlap in pairs of Lake Malawi cichlids (Martin and Genner, 2009). These comparative points would also be useful in assessing how patterns of resource use might alter in the future as it has been shown that climate change decreases productivity in LT (O'Reilly et al., 2003).

The isotopic samples in this study were mainly sampled from rocky littoral habitats (only *S. multipunctatus* was sampled from greater depths) which have been shown to host species which overlap in ecological niche space in other species, for example LT *Platythelphusa* crabs (Marijnissen et al., 2008), and Lake Malawi cichlids (Genner et al., 1999a). In order to collect enough samples to conduct the isotope analysis multiple collection methods were required and sampling focussed mainly on shallow areas, this has meant that it was not possible to investigate fine scale patterns including depth segregation which has been shown to segregate *Platythelphusa* crabs (Marijnissen et al., 2008) and has been hypothesised to lead to habitat segregation in LT *Chrysichthys* (Coulter, 1991). Many species inside and outside of LT have been shown to differentiate along a benthic/pelagic axis for example, East African cichlids (Cooper et al., 2010), notothenioid icefishes (Wilson et al., 2013), and multiple populations of three-spined sticklebacks (Willacker et al., 2010). Further samples from different depths in combination with samples different distances from the shore (as the extent that these taxa extend into LT is currently unknown) would enable this to be investigated more fully in the catfish radiations, especially as how habitat preferences change with ontogeny is currently unknown. Among the claroteine taxa that have been sampled at depth (*B. tetranema*, *C. sianenna* and LT *Chrysichthys*) *B. tetranema* and *C. sianenna,* which are the species with the most fusiform body shape, show the most pelagic $\delta^{13}C$ signatures (Figure 7; Figure 8),

however, more samples from the LT *Chrysichthys* which are found at greater depths (Coulter, 1991) would be needed to investigate if this was a simple depth pattern.

4.5.3 Morphological diversity and link with ecology

In the claroteine radiation the different genera were separated in morphospace in both analyses, though with more overlap in the non-phylogenetically corrected analysis, whereas both distantly and closely related taxa overlapped in isotopic niche space. This suggests that morphological similarity is not explained simply by recent ecological pressures but that phylogenetic history within the LT radiation is important. This pattern holds when looking at species with similar characteristics such as *L. brevispinis* and *Phyllonemus* sp. C both of which are small bodied, although they do overlap in isotopic niche (0.367 overlap in standard ellipse) they do not overlap in morphospace. When phylogeny is not taken into account the species with more fusiform body plans, *B. tetranema* and *C. sianenna*, overlap in morphospace but they do not occupy a similar area of morphospace in the phylogenetically corrected analysis.

In the non-phylogenetically corrected analysis the greatest area of morphospace was occupied by *L. asperispinis*, even though this was sampled from only four specimens including the holotype and paratopotype. The validity of this species has not been examined in a molecular phylogeny and it is only known from few specimens. This result may be due to this species displaying allometric effects that are not well known or alternatively the samples collected may include more than one species.

The *Synodontis* taxa are more similar morphologically and therefore their proximity in morphospace is closer than that seen in the claroteine radiation. In *Synodontis*, the main distinguishing characteristics (those corresponding to PC1) concern the size of the head (both height and length), size and position of the eye, snout length and body depth. The size of the caudal peduncle is among the characters associated with PC2. In the claroteines while head height and eye position are retained as distinguishing characters, gape width, body depth and the size of the caudal peduncle also correspond to PC1. Head length and eye diameter are among the PC2 features. These traits could potentially be involved in feeding

and broadly correspond to the differences seen in LT cichlids, for whom the relative size of the head and caudal peduncle are the main morphological differences observed (Clabaut et al., 2007). That differences in eye morphology are observed in both the diurnal *Synodontis* and the nocturnal claroteines is potentially interesting, given that the former is likely to be more reliant upon vision than the latter.

A high degree of overlap in morphology and ecology in phylogenetically distinct taxa is potentially suggestive of convergent evolution. This is not common in this study as the different claroteine genera cluster in morphospace. There is some evidence of convergence in the *Synodontis* between Clades 4 and 6 but these taxa are relatively closely related. In phylogenetically close taxa it is potentially representative of niche conservatism, a pattern possibly seen in *Lophiobagrus*, which display a large amount of overlap in the isotope analysis, and are relatively closely clustered in the morphology analysis, particularly on PC1.

In this study the correlation between morphology and ecology was not significant using phylogenetic generalized least squares regression or phylogenetic canonical correlation analysis. It is an important criterion of adaptive radiation that there is a correlation between divergent phenotypes and differentiation of ecological niches, however the lack of a significant correlation in this analysis does not necessarily rule out the possibility of an adaptive radiation. Because PC axes were used to distinguish morphology this simply means that the specific combination of traits associated with each PC axis (e.g., head height, eye position, gape width, body depth and the size of the caudal peduncle all correspond to PC1 in the claroteines) does not correlate with changes in either $\delta^{15}N$ or $\delta^{13}C$ signatures. In addition, in other lacustrine radiations, species feed opportunistically on a greater range of resources than those predicted by their specialised trophic morphologies (Genner et al., 1999b; Liem and Osse, 1975; Liem, 1980). This suggests that adaptive trophic specialisations can evolve without the exclusion of feeding generally on common resources. Future detailed geometric morphometric analyses, specifically of the head, may help to explicitly test if there is a correlation between trophic morphology and ecology.

4.5.4 Is there evidence that the radiations are adaptive?

In radiations that are adaptive, genetically distinct taxa derived from a single common ancestor should display morphological and physiological traits associated with the exploitation of differing resources. When visualised as disparity through time, this pattern would be expected to be reflected in a lower relative subclade disparity than expected under the null model if diversification occurred at the beginning of a radiation, if adaptive diversification occurred later then increased disparity within subclades may be present as niches were filled sequentially. However, a pattern of increased disparity within subclades compared to the entire radiation would also be seen if species from differing subclades independently evolved similar morphologies.

Synodontis show higher subclade disparity in morphology than expected under Brownian motion (the null model). This greater subclade diversification coupled with increased morphological similarities between subclades suggests that similarities have independently evolved in parallel. Initially subclade disparity in isotope signature is lower than the null model suggesting that adaptive niche filling may have played a role in this initial diversification. If niche conservatism, rather than divergence, is adaptive, a greater degree of ecological similarity than expected under Brownian motion is likely to be seen throughout the radiation. LT *Synodontis* species are known to be Müllerian mimics (Wright, 2011), with shared warning patterns, in this case the conservation of morphological forms may constitute a selective advantage. Constraints on morphology related to predator pressure may limit their ability to specialise with respect to feeding. The high levels of subclade disparity late in the radiation could be seen if more recent branching events were driven by other factors such as allopatric speciation coupled with selection to maintain morphological features or convergence in taxa that colonise the new environment (i.e. the southern basin) into forms already present in the older central and northern basins. In *Synodontis* a peak in subclade disparity occurs within the last 1 Ma. This peak is most significantly seen in PC1, and corresponds to the divergence of clades, 1, 4, and 5, all of which are relatively closely clustered on PC1. A similar, albeit less dramatic peak is seen in the claroteines, corresponding to diversification events in the LT *Chrysichthys* and slightly postdating branching events in *Lophiobagrus* and *Phyllonemus*.

In the claroteine radiation there is a decline in morphological subclade disparity that is faster than expected under a model of Brownian motion, especially at the beginning of the radiation which corresponds to the establishment of full lacustrine conditions (5-6Ma) (Tiercelin and Mondeguer, 1991). These negative MDI values (lower subclade disparity than expected under Brownian motion) are seen most strongly on PC1. These results are not significant, but they are potentially indicative of diversification into new niches (into new areas of morphospace) early in the radiation. Taxa that diversify early can do this by a process of niche filling in new environments with less ecological divergence later in the radiation. A similar pattern is seen when considering disparity in isotope signatures, with an initial drop in disparity with respect to Brownian motion, followed later by an increase in disparity above the null prediction. This may reflect early specialisation followed by later evolutionary constraint with respect to diet. This pattern of an initial drop in disparity is considered as indicative of adaptive radiation and is observed in radiations of body size in cetaceans (Slater et al., 2010) and Madagascan vangas (Jønsson et al., 2012).

The number of subclades increases through time, meaning that the number of taxa within each subclade decreases which increases the effect of incomplete taxonomic sampling and the chances of missing divergence that is currently in progress. In light of this, disparity through time analyses in older radiations can be truncated to include only a portion of the tree (≈ 66% of the tree) (e.g., López-Fernández et al., 2013). This is a substantial disadvantage for investigating patterns of diversification in recent and ongoing radiations, where a large proportion of diversification has taken place within the last 30% of the tree. This is potentially a problem in this study especially as both of the radiations investigated here are proposed to contain geographically restricted taxa and sampling is from throughout LT is limited. In particular, the lack of samples from the Democratic republic of Congo (DRC) may have an influence on the results of this analysis. The inclusion of DRC samples, if and when these become available, would potentially lend greater weight to the conclusions of this analysis. The inclusion of the putative species in the claroteine radiation and each identified mitochondrial clade of *Synodontis* mean that as much on-going divergence as possible was included in these analyses in order reduce the impact of these biases.

4.5.5 Conclusions

This study concludes that the LT *Synodontis* radiation is younger than the claroteine radiation and is more equivalent in age to the different genera within the claroteines. There is some evidence that the claroteine radiation is adaptive, specifically a lack of significant overlap in morphology coupled with a large initial drop in subclade disparity through time for both morphology and ecology. There is evidence of divergence in diet both between and within genera but there is also significant overlap in isotopic signatures between some species. Dietary divergence is most pronounced in the more fusiform species, which display signatures of a more pelagic diet. In contrast, the *Synodontis* radiation has disparity mainly partitioned within subclades, which could be the result of either recent diversification into different niches/areas of morphospace or convergence with species in disparate clades evolving similar phenotypes. The large overlap in stable isotope signatures and the evolution of a hind gut related to resource use in species that are not sister to each other provides some evidence of convergence. Although the overall patterns seen between the two radiations are dissimilar, the patterns seen within the individual genera of the claroteine radiation provide a useful comparative system to the *Synodontis* radiation.

4.6 References

Albertson, R. C. (2008) Morphological divergence predicts habitat partitioning in a Lake Malawi cichlid species complex. *Copeia*, **2008**(3), 689–698

Alexandrou, M. A., Oliveira, C., Maillard, M., McGill, R. A. R., Newton, J., Creer, S., & Taylor, M. I. (2011) Competition and phylogeny determine community structure in Müllerian co-mimics. *Nature*, **469**(7328), 84–88

Arif, S., Aguirre, W., & Bell, M. (2009) Evolutionary diversification of opercle shape in Cook Inlet threespine stickleback. *Biological Journal of the Linnean Society*, **97**(4), 832 - 844

Bailey, R. M., & Stewart, D. J. (1984) Bagrid catfishes from Lake Tanganyika, with a key and descriptions of new taxa. *Miscellaneous Publications Museum of Zoology, University of Michigan*, **168**

Barrow, L. M., Bjorndal, K. A., & Reich, K. J. (2008) Effects of preservation method on stable carbon and nitrogen isotope values. *Physiological and Biochemical Zoology*, **81**(5), 688–693

Biomatters (2013) Geneious version 5.6 created by Biomatters. <http://www.geneious.com>

Blomberg, S. P., Garland, T., & Ives, A. R. (2003) Testing for phylogenetic signal in comparative data: behavioral traits are more labile. *Evolution*, **57**(4), 717–745

Brown, K. J., Rüber, L., Bills, R., & Day, J. J. (2010) Mastacembelid eels support Lake Tanganyika as an evolutionary hotspot of diversification. *BMC Evolutionary Biology*, **10**, 188

Bruner, J. (1991) Comments on the genus Amyzon (Family Catostomidae). *Journal of Paleontology*, **65**(4), 678–686

Clabaut, C., Bunje, P. M. E., Salzburger, W., & Meyer, A. (2007) Geometric morphometric analyses provide evidence for the adaptive character of the Tanganyikan cichlid fish radiations. *Evolution*, **61**(3), 560–578

Cooper, W. J., Parsons, K., McIntyre, A., Kern, B., McGee-Moore, A., & Albertson, R. C. (2010) Bentho-pelagic divergence of cichlid feeding architecture was prodigious and consistent during multiple adaptive radiations within African rift-lakes. *PloS One*, **5**(3), e9551

Correa, C. (2012) Tissue preservation biases in stable isotopes of fishes and molluscs from Patagonian lakes. *Journal of Fish Biology*, **81**(6), 2064–2073

Coulter, G. W. (1991) The benthic fish communities In G. W. Coulter (Ed.), *Lake Tanganyika and its Life.* Oxford: Oxford University Press. pp. 151–199

Coulter, G. W., & Spigel, R. H. (1991) Hydrodynamics, In G. W. Coulter (Ed.), *Lake Tanganyika and its Life.* Oxford: Oxford University Press. pp. 49–75

Davis, M. P., Arratia, G., & Kaiser, T. M. (2013) The first fossil shellear and its implications for the evolution and divergence of the Kneriidae (Teleostei: Gonorynchiformes), In Arriata, G., Schultze, H.-P. & Wilson, M. V. H. (Eds.), *Mesozoic Fishes* 5-*Global Diversity and Evolution*. pp. 325–362

Day, J. J., Bills, R., & Friel, J. P. (2009) Lacustrine radiations in African Synodontis catfish. *Journal of Evolutionary Biology*, **22**(4), 805–817

Day, J. J., Peart, C. R., Brown, K. J., Friel, J. P., Bills, R., & Moritz, T. (2013) Continental diversification of an African catfish radiation (Mochokidae: Synodontis) *Systematic Biology*, **62**(3), 351-365

Day, J. J., & Wilkinson, M. (2006) On the origin of the Synodontis catfish species flock from Lake Tanganyika. *Biology Letters*, **2**(4), 548–552

Drummond, A. J., Suchard, M. A., Xie, D., & Rambaut, A. (2012) Bayesian phylogenetics with BEAUti and the BEAST 1.7. *Molecular Biology and Evolution*, **29**(8), 1969–1973

Dufour, E., & Gerdeaux, D. (2001) Contribution of stable isotopes to fish ecological studies. *Cybium*, **25**(4), 369–382

Freckleton, R. P., & Harvey, P. H. (2006) Detecting non-Brownian trait evolution in adaptive radiations. *PLoS Biology*, **4**(11), e373

Fryer, G., & Iles, T. D. (1972) *The cichlid fishes of the Great Lakes of Africa*. Edinburgh: Oliver and Boyd

Genner, M. J., & Turner, G. F. (2012) Ancient hybridization and phenotypic novelty within Lake Malawi's cichlid fish radiation. *Molecular Biology and Evolution*, **29**(1), 195–206

Genner, M. J., Turner, G. F., & Hawkins, S. J. (1999a) Foraging of rocky habitat cichlid fishes in Lake Malawi: coexistence through niche partitioning? *Oecologia*, **121**, 283–292.

Genner, M. J., Turner, G. F., Barker, S., & Hawkins, S. J. (1999b) Niche segregation among Lake Malawi cichlid fishes? Evidence from stable isotope signatures. *Ecology Letters*, **2**, 185–190

Greenwood, P. H., & Patterson, C. (1967) A fossil osteoglossoid fish from Tanzania (E . Africa). *Zoological Journal of the Linnean Society*, **47**(311), 211–223

Hammer, Ø., Harper, D. A. T., & Ryan, P. D. (2001) PAST: Paleontological Statistics software package for education and data analysis. *Palaeontologica Electronica*, **4**(1), 9

Hardman, M. (2008) A new species of catfish genus Chrysichthys from Lake Tanganyika (Siluriformes: Claroteidae). *Copeia*, **2008**(1), 43

Harmon, L. J., Schulte, J. A, Larson, A., & Losos, J. B. (2003) Tempo and mode of evolutionary radiation in iguanian lizards. *Science ,* **301**(5635), 961–964

Harmon, L. J., Weir, J. T., Brock, C. D., Glor, R. E., & Challenger, W. (2008) GEIGER: investigating evolutionary radiations. *Bioinformatics*, **24**(1), 129–131

Harrison, T., Msuya, P., Murray, A., Jacobs, B. F., Ba'ez, A. M., Mundil, R., & Ludwig, K. R. (2001) Paleontological investigations at the Eocene locality of Mahenge in north-central Tanzania, East Africa. In G. F. Gunnell (Ed.), *Eocene biodiversity: unusual and rarely sampled habitats.* New York: Kluwer Academic/Plenum Publishers. pp. 39–74

Hesslein, R. H., Hallard, K. A., & Ramlal, P. (1993) Replacement of sulfur, carbon, and nitrogen in tissue of growing broad whitefish (Coregonus nasus) in response to a change in diet traced by delta-S-34, delta-C-13 and delta-N-15. *Canadian Journal of Fisheries and Aquatic Sciences*, **50**(10), 2071–2076

Hobson, K. A., Gibbs, H. L., & Gloutney, M. L. (1997) Preservation of blood and tissue samples for stable-carbon and stable-nitrogen isotope analysis. *Canadian Journal of Zoology*, **75**, 1720–1723

Hori, M., Gashagaza, M. M., Nshombo, M., & Kawanabe, H. (1993) Littoral fish communities in Lake Tanganyika: irreplaceable diversity supported by intricate interactions among species. *Conservation Biology*, **7**, 657–666

Jønsson, K. A, Fabre, P.-H., Fritz, S. a, Etienne, R. S., Ricklefs, R. E., Jørgensen, T. B., & Irestedt, M. (2012) Ecological and evolutionary determinants for the adaptive radiation of the Madagascan vangas. *Proceedings of the National Academy of Sciences*, **109**(17), 6620–6625

Keller, I., Wagner, C. E., Greuter, L., Mwaiko, S., Selz, O. M., Sivasundar, A., & Seehausen, O. (2013) Population genomic signatures of divergent adaptation, gene flow and hybrid speciation in the rapid radiation of Lake Victoria cichlid fishes. *Molecular Ecology*, **22**(11), 2848–2863

Kelly, B., Dempson, J. B., & Power, M. (2006) The effects of preservation on fish tissue stable isotope signatures. *Journal of Fish Biology*, **69**(6), 1595–1611

Kembel, S. W., Cowan, P. D., Helmus, M. R., Cornwell, W. K., Morlon, H., Ackerly, D. D., & Webb, C. O. (2010) Picante: R tools for integrating phylogenies and ecology. *Bioinformatics*, **26**(11), 1463–1464

Kiljunen, M., Grey, J., Sinisalo, T., Harrod, C., Immonen, H., & Jones, R. I. (2006) A revised model for lipid-normalizing δ13C values from aquatic organisms, with implications for isotope mixing models. *Journal of Applied Ecology*, **43**(6), 1213–1222

Koblmüller, S., Salzburger, W., Obermüller, B., Eigner, E., Sturmbauer, C., & Sefc, K. M. (2011) Separated by sand, fused by dropping water: habitat barriers and fluctuating water levels steer the evolution of rock-dwelling cichlid populations in Lake Tanganyika. *Molecular Ecology*, **20**(11), 2272–2290

Koblmüller, S., Sturmbauer, C., Verheyen, E., Meyer, A., & Salzburger, W. (2006) Mitochondrial phylogeny and phylogeography of East African squeaker catfishes (Siluriformes: Synodontis). *BMC Evolutionary Biology*, **6**, 49

Kozak, K. H., Weisrock, D. W., & Larson, A. (2006) Rapid lineage accumulation in a non-adaptive radiation: phylogenetic analysis of diversification rates in eastern North American woodland salamanders (Plethodontidae: Plethodon). *Proceedings of the Royal Society B: Biological Sciences*, **273**(1586), 539–446

Kozak, K. H., & Wiens, J. J. (2006) Does niche conservatism promote speciation? A case study in North American salamanders. *Evolution*, **60**(12), 2604–2621

Lanfear, R., Calcott, B., Ho, S. Y. W., & Guindon, S. (2012) Partitionfinder: combined selection of partitioning schemes and substitution models for phylogenetic analyses. *Molecular Biology and Evolution*, **29**(6), 1695–1701

Lavoué, S., Miya, M., Inoue, J. G., Saitoh, K., Ishiguro, N. B., & Nishida, M. (2005) Molecular systematics of the gonorynchiform fishes (Teleostei) based on whole mitogenome sequences: implications for higher-level relationships within the Otocephala. *Molecular Phylogenetics and Evolution*, **37**(1), 165–177

Li, C., Ortí, G., Zhang, G., & Lu, G. (2007) A practical approach to phylogenomics: the phylogeny of ray-finned fish (Actinopterygii) as a case study. *BMC Evolutionary Biology*, **7**(1), 44

Liem, K. (1980) Adaptive significance of intra-and interspecific differences in the feeding repertoires of cichlid fishes. *American Zoologist*, **20**(1), 295–314

Liem, K. F., & Osse, J. W. M. (1975) Biological versatility, evolution, and food resource exploitation in African cichlid fishes. *American Zoologist*, **454**, 427–454

López-Fernández, H., Arbour, J. H., Winemiller, K. O., & Honeycutt, R. L. (2013) Testing for ancient adaptive radiations in neotropical cichlid fishes. *Evolution*, **67**(5), 1321–1337

Lundberg, J. G. (1975) The fossil catfishes of North America. Claude W. Hibbard Memorial, Volume 2. *University of Michigan, Papers on Paleontology*, **11**, 1–51

Lundberg, J. G. (1992) The phylogeny of ictalurid catfishes: A synthesis of recent work. In R. Mayden (Ed.), *Systematics, Historical Ecology, and North American Freshwater Fishes.* Stanford :Stanford University Press. pp. 392–420

Lundberg, J. G., Sullivan, J. P., Rodiles-hernández, R., & Dean, A. (2007) Discovery of African roots for the Mesoamerican Chiapas catfish , Lacantunia enigmatica , requires an ancient intercontinental passage, **156**(1), 39–53

Marijnissen, S. A. E., Michel, E., Kamermans, M., Olaya-Bosch, K., Kars, M., Cleary, D. F. R., & Menken, S. B. J. (2008) Ecological correlates of species differences in the Lake Tanganyika crab radiation. *Hydrobiologia*, **615**(1), 81–94

Martin, C. H., & Genner, M. J. (2009) High niche overlap between two successfully coexisting pairs of Lake Malawi cichlid fishes. *Canadian Journal of Fisheries and Aquatic Sciences*, **66**(4), 579–588

Montaña, C. G., & Winemiller, K. O. (2013) Evolutionary convergence in Neotropical cichlids and Nearctic centrarchids: evidence from morphology, diet, and stable isotope analysis. *Biological Journal of the Linnean Society*, **109**(1), 146–164

Muschick, M., Indermaur, A., & Salzburger, W. (2012) Convergent evolution within an adaptive radiation of cichlid fishes. *Current Biology*, **22**(24), 2362–2368

Naithani, J., Deleersnijder, E., & Plisnier, P. (2003) Analysis of wind-induced thermocline oscillations of Lake Tanganyika. *Environmental Fluid Mechanics*, **3**(1), 23–39

Nakatani, M., Miya, M., Mabuchi, K., Saitoh, K., & Nishida, M. (2011) Evolutionary history of Otophysi (Teleostei), a major clade of the modern freshwater fishes: Pangaean origin and Mesozoic radiation. *BMC Evolutionary Biology*, **11**(1), 177

Near, T. J., Eytan, R. I., Dornburg, A., Kuhn, K. L., Moore, J. A, Davis, M. P., & Smith, W. L. (2012) Resolution of ray-finned fish phylogeny and timing of diversification. *Proceedings of the National Academy of Sciences* , **109**(34), 13698–13703

O'Reilly, C. M., Alin, S. R., Plisnier, P.-D., Cohen, A. S., & McKee, B. A. (2003) Climate change decreases aquatic ecosystem productivity of Lake Tanganyika, Africa. *Nature*, **424**(6950), 766–768

O'Reilly, C. M., Hecky, R. E., Cohen, A. S., & Plisnier, P.-D. (2002) Interpreting stable isotopes in food webs: recognizing the role of time averaging at different trophic levels. *Limnology and Oceanography*, **47**(1), 306–309

Ochi, H., Rossiter, A., & Yanagisawa. (2000) The first record of a biparental mouthbrooding catfish. *Journal of Fish Biology*, **57**(6), 1601–1604

Ochi, H., Rossiter, A., & Yanagisawa, Y. (2001) Biparental mouthbrooding of the catfish Phyllonemus filinemus in Lake Tanganyika. *Ichthyological Research*, **48**, 225–229

Ochi, H., Rossiter, A., & Yanagisawa, Y. (2002) Paternal mouthbrooding bagrid catfishes in Lake Tanganyika. *Ichthyological Research*, **49**, 270–273

Ogawa, N. O., Koitabashi, T., Oda, H., Nakamura, T., & Wada, E. (2001) Fluctuations of nitrogen isotope ratio of gobiid fish (Isaza) specimens and sediments in Lake Biwa, Japan, during the 20th century. *Limnology and Oceanography*, **46**(5), 1228–1236

Parnell, A. C., Inger, R., Bearhop, S., & Jackson, A. L. (2010) Source partitioning using stable isotopes: coping with too much variation. *PloS One*, **5**(3), e9672

Peart, C. R., Bills, R., Wilkinson, M., & Day, J. J. (2014) Nocturnal claroteine catfishes reveal dual colonisation but a single radiation in Lake Tanganyika. *Molecular Phylogenetics and Evolution*, **73**, 119–128

Pinheiro, J., Bates, D., DebRoy, S., Sarkar, D., & R Core Team. (2014) nlme: Linear and Nonlinear Mixed Effects Models. R package version 3.1-117 <http://CRAN.R-project.org/package=nlme>

Plisnier, P.-D., Chitamwebwa, D., Mwape, L., Tshibangu, K., Langenberg, V., & Coenen, E. (1999) Limnological annual cycle inferred from physical-chemical fluctuations at three stations of Lake Tanganyika. *Hydrobiologia*, **407**, 45–58

Polly, P. D., Lawing, A. M., Fabre, A.-C., & Goswami, A. (2013) Phylogenetic principal components analysis and geometric morphometrics. *Hystrix,* **24**(1), 33–41

Post, D. (2002) Using stable isotopes to estimate trophic position : models, methods, and assumptions. *Ecology*, **83**(3), 703–718

Poyato-Ariza. (1996) A revision of Rubiesichthys gregalis WENZ 1984 (Ostariophysi, Gonorynchiformes), from the Early Cretaceous of Spain. In G. Arratia & G. Viohl (Eds.), *Mesozoic Fishes: Systematics and Paleoecology.* Munich: Dr. Friedrich Pfeil. pp. 319–328

Poyato-Ariza, F. J., Grande, T., & Diogo, R. (2010) Gonorynchiform interrelationships: Historic Overview, analysis, and revised systematics of the group. In T. Grande, F. J. Poyato-Ariza, & R. Diogo (Eds.), *Gonorynchiformes and Ostariophysan Relationships: A Comprehensive Review.* Enfield: Science Publishers. pp. 227–338

Rambaut, A., & Drummond, A. J. (2009) Tracer v1.5 <http://beast.bio.ed.ac.uk/Tracer>

Revell, L. J. (2009) Size-correction and principal components for interspecific comparative studies. *Evolution*, **63**(12), 3258–3268

Rüber, L., & Adams, D. (2001) Evolutionary convergence of body shape and trophic morphology in cichlids from Lake Tanganyika. *Journal of Evolutionary Biology*, **14**, 325–332

Rüber, L., Verheyen, E., & Meyer, A. (1999) Replicated evolution of trophic specializations in an endemic cichlid fish lineage from Lake Tanganyika. *Proceedings of the National Academy of Sciences*, **96**(18), 10230–10235

Salzburger, W., Baric, S., & Sturmbauer, C. (2002) Speciation via introgressive hybridization in East African cichlids? *Molecular Ecology*, **11**(3), 619–625

Sarakinos, H., Johnson, M. L., & Vander Zanden, M. J. (2002) A synthesis of tissue-preservation effects on carbon and nitrogen stable isotope signatures. *Canadian Journal of Zoology*, **80**(2), 381–387

Sato, T. (1986) A brood parasitic catfish of mouthbrooding cichlid fishes in Lake Tanganyika. *Nature*, **323**, 58–59

Schluter, D. (1996) Ecological Speciation in Postglacial Fishes. *Philosophical Transactions of the Royal Society B: Biological Sciences*, **351**(1341), 807–814

Schluter, D. (2000) *The Ecology of Adaptive Radiation*. Oxford: Oxford University Press.

Seehausen, O., Terai, Y., Magalhaes, I. S., Carleton, K. L., Mrosso, H. D. J., Miyagi, R., & Okada, N. (2008) Speciation through sensory drive in cichlid fish. *Nature*, **455**(7213), 620–626

Slater, G. J., Price, S. A, Santini, F., & Alfaro, M. E. (2010) Diversity versus disparity and the radiation of modern cetaceans. *Proceedings of the Royal Society B: Biological Sciences*, **277**(1697), 3097–3104

Smith, G. R. (1992) Phylogeny and biogeography of the Catostomidae, freshwater fishes of North America and Asia. In R. Mayden (Ed.), *Systematics, Historical Ecology, and North American Freshwater Fishes.* Stanford: Stanford University Press. pp. 778–826

Sturmbauer, C., Mark, W., & Dallinger, R. (1992) Ecophysiology of Aufwuchs-eating cichlids in Lake Tanganyika: niche separation by trophic specialization. *Environmental Biology of Fishes*, **35**, 283–290

Sullivan, J. P., Lundberg, J. G., & Hardman, M. (2006) A phylogenetic analysis of the major groups of catfishes (Teleostei: Siluriformes) using rag1 and rag2 nuclear gene sequences, **41**(3), 636–662

Tiercelin, J. J., & Mondeguer, A. (1991) The geology of the Tanganyika trough. In G. W. Coulter (Ed.), *Lake Tanganyika and its Life.* Oxford: Oxford University Press. pp. 7–48

Verheyen, E., Rüber, L., Snoeks, J., & Meyer, a. (1996) Mitochondrial phylogeography of rock-dwelling cichlid fishes reveals evolutionary influence of historical lake level fluctuations of Lake Tanganyika, Africa. *Philosophical Transactions of the Royal Society B: Biological Sciences*, **351**(1341), 797–805

Vigliotta, T. (2008) A phylogenetic study of the African catfish family Mochokidae (Osteichthyes, Ostariophysi, Siluriformes), with a key to genera. *Proceedings of the Academy of Natural Sciences of Philadelphia,* **157**(1), 73–136

Willacker, J. J., von Hippel, F. A, Wilton, P. R., & Walton, K. M. (2010) Classification of threespine stickleback along the benthic-limnetic axis. *Biological Journal of the Linnean Society.*, **101**(3), 595–608

Wilson, L. A. B., Colombo, M., Hanel, R., Salzburger, W., & Sánchez-Villagra, M. R. (2013) Ecomorphological disparity in an adaptive radiation: opercular bone shape and stable isotopes in Antarctic icefishes. *Ecology and Evolution*, **3**(9), 3166–3182

Wilson, M. (1993) Calibration of Eocene varves at Horsefly, British Columbia, Canada, and temporal distribution of specimens of the Eocene fish Amyzon aggregatum. *Kaupia Darmstaedter Beitraege Zur Naturgeschichte*, **2**, 27–38

Winkelmann, K., Genner, M. J., Takahashi, T., & Rüber, L. (2014) Competition-driven speciation in cichlid fish. *Nature Communications*, **5**, 3412

Wright, J. J. (2011) Conservative coevolution of Müllerian mimicry in a group of rift lake catfish. *Evolution*, **65**(2), 395–407

Wright, J. J., & Bailey, R. M. (2012) Systematic revision of the formerly monotypic genus Tanganikallabes (Siluriformes: Clariidae). *Zoological Journal of the Linnean Society*, **165**(1), 121–142

Wright, J. J., & Page, L. M. (2006) Taxonomic revision of Lake Tanganyikan Synodontis (Siluriformes: Mochokidae). *Bulletin of the Florida Museum of Natural History*, **46**(4), 99–154

Xu, J., Yang, Q., Zhang, M., Zhang, M., Xie, P., & Hansson, L.-A. (2011) Preservation effects on stable isotope ratios and consequences for the reconstruction of energetic pathways. *Aquatic Ecology*, **45**(4), 483–492

Yuma, M., Narita, T., Hori, M., & Kondo, T. (1998) Food resources of shrimp-eating cichlid fishes in Lake Tanganyika. *Environmental Biology of Fishes*, **52**, 371–378.

Chapter 5
Morphological divergence at a continental scale in *Synodontis* catfishes

5.1 Abstract

The study of evolutionary radiations is vital to understanding the patterns and processes leading to species diversity, however the majority of studies to date have focused on insular radiations. While significant inferences have been made, it is important to evaluate whether the same factors influence diversification at a continental scale. The catfish genus *Synodontis* offers a valuable model in which to investigate these patterns, as although it is mainly riverine with a broad sub-Saharan distribution, it also contains a small radiation (≈11 species) in Lake Tanganyika. In this study morphological diversification across the genus is investigated using morphological measurements from 612 specimens and morphological convergence between different river basins was investigated using the R package SURFACE. There is considerable overlap in morphospace of species from different geographic areas and disparity is mainly partitioned within subclades, suggesting that the independent evolution of taxa in different drainage basins produced similar morphologies. There is little evidence of explicit convergence detected by SURFACE within the genus, however, suggesting that this pattern may be driven by phenotypic constraints. This is also evidenced by the Lake Tanganyika taxa which are not found to be morphologically distinct from their riverine congeners, despite their relatively recent lacustrine radiation. Incomplete sampling from throughout Africa is a problem in continental scale studies and this potential bias is investigated in *Synodontis* using geo-referenced museum collections (6,194 specimens). Sampling for this genus is broad and though some areas are identified as priorities for future collecting (e.g., Angola), *Synodontis* provide a useful system in which to investigate the patterns driving diversification at a continental scale in freshwater environments.

5.2 Introduction

Key to understanding the diversity of species, and the differences and similarities between them, is the study of the patterns and processes driving diversification. In general, however, research in this area has tended to focus on the 'natural experiments' of island or lake radiations. These studies have led to some major ideas, for example divergence by habitat, then trophic partitioning followed by further diversification based on behaviour or sexual selection (e.g., Lake Malawi cichlids, Danley and Kocher, 2001) as a potential mode for adaptive radiation. This tends to reveal itself by major morphological changes in shape related to habitat shift or trophic adaptations earlier in a radiation followed by smaller changes, for example in colour pattern, later. These radiations have been used to search for general patterns, for example whether similar convergent phenotypes can co-exist as seen in Lake Tanganyika (LT) cichlids (Muschick et al., 2012) if the radiation is repeatable, for example the same phenotypes evolve repeatedly in East African cichlids (Cooper et al., 2010; Fryer and Iles, 1972), or if radiations of entire faunas occur in parallel, as observed in Caribbean *Anolis* lizards (Mahler et al., 2013). However, questions remain as to how representative these well studied systems are of the majority of non-marine taxa which live on continents, particularly as, with some notable exceptions (e.g., East African cichlids), 'island' radiations tend to contain fewer species than their continental counterparts.

It is not clear if 'island' and continental radiations follow the same patterns of diversification, in terms of either lineage formation or phenotypic divergence, and even the same patterns are no guarantee that they result from the same evolutionary processes. For example, the model of adaptive radiation mentioned above shows early rapid cladogenesis followed by a decline, but this same pattern can be seen as a result of geographic speciation (Pigot et al., 2010), which might be expected to be more likely on the larger continents, as isolation is common and often linked to changing geology and climate. A small, recent continental radiation which shows a similar pattern to that seen in island radiations is the capuchino seedeaters of South America (11 species). These birds show clear phenotypic variation in male plumage colour and pattern, but very little genetic variation (Campagna et al., 2012).

In general, patterns of morphological diversification at a continental scale are not clear and may be more complicated than those seen on smaller spatial

scales. Dispersal limitation is thought to be a major factor rather than interactions between species, for example, in passerine birds morphological volume is unrelated to the number of species in a region (Ricklefs, 2012). There is also evidence that continental radiations may not be as constrained by ecological opportunities as island radiations in the Furnariidae (neotropical birds), which show a constant rate of diversification with evidence of constrained morphological evolution (Derryberry et al., 2011). Limited adaptive disparity is also seen in woodland salamanders found over a large area in North America, however, this clade shows rapid lineage accumulation, thought to be driven by phylogenetic niche conservatism and vicariant isolations forming strong barriers to geographic spread (Kozak et al., 2006).

Continental scale radiations also offer a good phylogenetic context within which to study island radiations, allowing the difference between these geographically restricted radiation and their continental relatives to be elucidated. For example, when the continental relatives of the Hawaiian honeycreepers (which have radiated in the Hawaiian archipelago) are compared to the continental relatives of the Hawaiian thrushes (which have not radiated on the islands), they display significantly greater variation in bill morphology, suggesting that island radiations just represent extreme versions of a clade specific ability to evolve these novel morphologies (Lovette et al., 2002). However, relationships between morphology and ecology can differ between continental and island environments which may make morphological patterns difficult to compare (Irschick et al., 1997).

Freshwater environments represent good systems in which to study continental diversification patterns, especially due to their restricted distributions and sensitivity to climatic changes. The hydrogeological hypothesis has been put forward to explain diversification patterns in Neotropical fish groups (Montoya-Burgos, 2003) suggesting that they were driven by palaeohydrological changes promoting dispersal and vicariance. However, this hypothesis has yet to be extensively tested in African taxa (with the exception of *Synodontis,* Day et al., 2013) and how this pattern of diversification affects morphological changes is not well understood. Even in the well-studied cichlids, while there are some densely sampled African phylogenies (e.g., Wagner et al., 2012), the patterns of morphological diversification outside of lake systems remain to be studied.

The endemic African catfish genus *Synodontis* has a broad distribution throughout sub-Saharan Africa and the Nile valley. The clade is species rich with 129 species (Vreven and Zamba, 2010) that are predominantly riverine, occurring mainly in mature rivers, though lower diversity is also found in the East African rift lakes (Poll, 1971) including ≈ 11 species in LT. The Mochokidae, of which *Synodontis* comprise around 65% of species diversity, were found to have high rates of species diversification and body size evolution, based on a time calibrated phylogeny across all fish (Rabosky et al., 2013). However, high rates of cladogenesis were not found when *Synodontis* was considered at the species level (Day et al., 2013) suggesting that macroevolutionary studies may not have the power to reveal patterns at low taxonomic scales. To date morphological diversification has not been investigated at the species level.

Synodontis are the only non-cichlid continental African radiation for which a molecular multi-gene phylogeny is available (Day et al., 2013) containing ≈ 60% of the currently described species. This study suggests that river palaeohydrology has been important in diversification, and although peak diversity is in the Congo basin, *Synodontis* are reconstructed to have originated and diversified in the West Africa-Nilo-Sudan region (broadly defined, see map in Figure 1 and Figure 2). The most recently colonised area is in Southern Africa, the Zambezi ichthyoprovince, which contains polyphyletic taxa indicating multiple independent colonisations.

In contrast to many island radiations *Synodontis* show a constant rate of lineage accumulation from mid-Cenozoic to the present (Day et al., 2013). This is despite diversifying during times of major climatic and geological events including East African rifting (10-25 Ma)(Partridge et al., 1995) and the middle Miocene climatic optimum (15-17 Ma), a warm period with high precipitation (Zachos, et al., 2001) leading to high connectivity between rivers. The dating estimates of the Day et al., (2013) study are based on a single *Synodontis* fossil calibration known from fin spines and a maximum hard bound on the root, and are different from those in the phylogenies generated in Chapter 4, which have significantly reduced taxonomic sampling within *Synodontis* but greater sampling outside and multiple fossil calibrations, and give younger ages. This is evidenced by the age of the most recent common ancestor of *Mochokus niloticus* and *Synodontis.* In the Day et al., (2013) phylogeny this is 48.48 Ma (41.43-58.3 Ma 95% Highest Posterior Density (HPD)), while in Chapter 4, in the Ostariophysian phylogeny this node is dated at

19.11 Ma (11.2-26.02 Ma HPD) and in the secondarily calibrated 'Big Africa' phylogeny at 24.05 Ma (16.96-32.02 Ma HPD). These alternate estimates indicate discordance between different dating methodologies and calibrations. It is problematic to assign fossils to the genus *Synodontis* as it is defined using two soft tissue character synapomorphies (Vigliotta, 2008). Nevertheless, regardless of the dating methodology, *Synodontis* is still diversifying over a large time span affected by major climatic and geological events.

The rates of species diversification in *Synodontis* are not related to species richness (Day et al., 2013) but the fastest rates are found in the East African clade which contains the LT radiation. There have been several studies into the evolutionary relationships in the LT taxa (Day and Wilkinson, 2006; Day et al., 2009; Koblmüller et al., 2006) but the factors driving diversification remain to be elucidated. Chapter 4 showed that LT *Synodontis* have wide niches with considerable overlap and disparity through time plots showed that relative subclade disparity in morphology was higher than expected under Brownian motion (suggesting a spread of values within subclades making them more similar to other subclades), perhaps providing an indication of convergence within the LT clade. These species have also been hypothesised to be Müllerian mimics (Wright, 2011) which may provide a selective pressure to be more similar to each other morphologically, however the presence of convergence in *Synodontis* has not been explicitly tested prior to this study.

This chapter addresses the patterns of morphological diversity and disparity across the entire *Synodontis* radiation for the first time and investigates if morphological convergence occurs in this group. In addition, the geographic spread and depth of museum collections is investigated in order to investigate the potential for sampling bias to influence these conclusions. The multi-gene phylogeny of Day et al., (2013) was used in combination with newly collected morphological measurements and geo-referenced museum collections to address the following questions; i) Do the LT taxa differ morphologically from the riverine *Synodontis* species? ii) Does morphological change reflect the constant rate pattern of lineage accumulation? iii) Is there convergent evolution of phenotypes between different drainage basins or climatic zones? iv) Is evidence of the recent colonisation of Southern Africa reflected in changes in morphology? v) Is there any evidence of geographic bias in sampling of *Synodontis* at a continental level?

5.3 Methods

5.3.1 Tree preparations

The fossil calibrated maximum clade credibility (MCC) phylogeny of Day et al., (2013) was used for this study after the outgroups had been pruned. This tree was generated using a matrix of 3586 base pairs including mitochondrial (*Cytb* + tRNA-pro and CO1) and nuclear (S7 intron 1 and intron 2 and RAG2) markers. This phylogeny was calibrated using a *Synodontis* fossil applied to the *Synodontis* crown group from the early Lower Oligocene dated to 34 Ma (Otero and Gayet, 2001; Roger et al., 1993). The age estimates for *Synodontis* and its constituent clades from this phylogeny are not congruent with the dates from the Ostariophysian and 'Big Africa' phylogenies from Chapter 4, however the phylogeny from Day et al., 2013 is preferred for the analyses in this chapter as it has the most complete taxonomic sampling.

The phylogeny was further pruned to include only one specimen per species and to remove taxa for which morphological data was not available. Some of the remaining specimens used in the phylogeny showed an affinity with a named species suggesting that they were closely related but not the same species. In these cases (*Brachysynodontis* aff *batensoda*, *S.* aff *bastiani, S*. aff *laessoei*) morphology measurements of the named taxa were used as the specimens in the phylogeny are thought to be closely related to the named species, however this may be a potential source of error. *Synodontis wamiensis* was also pruned as, following the taxonomy of Catalog of fishes (Eschmeyer, 2014), this is a synonym of *S. rukwaensis*. The same morphological data as used in Chapter 4 was used for the Lake Tanganyika taxa. Each of the unnamed clades from Chapter 4 is included in the published phylogeny of Day et al. (2013), but with an assigned name that may not be correct, (as evidenced by the taxonomic uncertainty (aff) and discussed more fully in Chapter 4). The 'Clade' nomenclature is retained in this Chapter. *Synodontis grandiops* was also pruned from the phylogeny as this specimen (CU91902) resolves within the *S. multipunctatus* samples (discussed in Chapter 4) including specimens also used in Chapter 3. Some of the undescribed specimens from the Day et al. (2013) phylogeny were able to be included as the specimens were available to be measured (*Synodontis* sp. nov. 2, *Synodontis* sp. nov. 3, *Synodontis* sp. nov. 4).

5.3.2 Principal components analysis

This study uses standard length in addition to the morphological measurements used in Chapter 4 (head length from the upper lip to the posterior edge of the operculum, the eye diameter, head height through the centre of the eye, eye position measured as the distance between the base of the head to the centre of the eye, snout length from the upper lip to the centre of the eye, maximum body depth, depth of the caudal peduncle and body width at pectoral fin inserts) with the exception of gape width. Some *Synodontis* species have very dense mandibular teeth, usually for scraping algae, making it difficult to measure gape width, so snout width, an alternate measure of how wide the head is at the mouth was used, defined as the width of the snout at the anterior nostrils.

Both specimens of species that were included in the phylogeny (394 specimens from 70 species, minimum 1, maximum 14, median 5 specimens per species) and not included in the phylogeny (218 specimens from 53 species, minimum 1, maximum 10, median 5 specimens per species) were measured for morphological analyses. 313 specimens from the Royal Museum of Central Africa, 156 specimens from the Natural History Museum, London, 53 specimens from the Cornell Museum of Vertebrates, 31 specimens from the American Museum of Natural History and 59 specimens collected directly were included in the dataset, a total of 612 individuals.

The morphological measurements used in this study were taken by two investigators (Claire Peart and Julia Day), a potential source of bias, however, a pilot study was conducted and inter-investigator error in measurements was not found to be different to the low level of intra-investigator error when measuring the same specimen multiple times. The natural logarithm of each measurement was taken before the measurements were averaged by species. The (logged and averaged) standard length was used to size correct the remaining morphological measurements taking phylogeny into account in the R package phytools (Revell, 2009). The residuals from this analysis were used to conduct a phylogenetic PCA in the same R package. A phylogenetically controlled approach was used to take into account the non-independence of trait values due to their relatedness, however, it has been suggested that these results can be difficult to interpret (Polly et al., 2013) and the phylogenetic PCA explicitly uses an evolutionary model of Brownian motion which does not necessarily fit the data. In order to assess the

contribution of each trait to the principal component loadings, the analysis was also repeated without taking phylogeny into account in PAST (Hammer et al., 2001). In this analysis the measurements were size corrected using the allometric vs standard option. All of the species, including those not in the molecular phylogeny, were size-corrected and a PCA performed in PAST as described above in order to investigate if these additional species occupy different areas of morphospace.

In order to visualise the results of the phylogenetic PCA the 'morphospace wheel plot' of Grundler and Rabosky (2014) was used in order to view three-dimensional morphospace in two dimensions and emphasise the co-variation of points. This method uses the same measures used to describe a point in 3-D morphospace, its angle in the xy plane, angle above the xy plane and its distance from the origin and represents them in two dimensions. This is achieved through the use of polar co-ordinates, the polar coordinate is assigned the angle in the xy plane, the radial coordinate is equal to the angle above the xy plane and the size of the point is proportional to the distance from the origin.

The command fitcontinuous in the R package GEIGER (Harmon et al., 2008) was used to assess the fit of the Brownian motion, Ornstein-Uhlenbeck (OU) and Early Burst models of continuous character evolution to each retained PC axes.

5.3.3 Disparity through time

The patterns of morphological diversification through time (in terms of each PC axis separately and all of the retained axes combined) were visualised using disparity through time plots in the R package GEIGER. This method (described in Chapter 4) shows the mean relative disparity at each node (disparity in a subclade divided by total disparity in the clade) of all subclades that had ancestral lineages present at that point in time. On each plot 10,000 simulations under the null model of Brownian motion were completed using the MCC tree with 95% confidence levels computed as in Slater et al., (2010). The morphological disparity index (MDI) was also computed (Harmon et al., 2003; described in Chapter 4) in order to assess if subclade disparity is higher or lower than expected under the null model. This was also repeated using a subset of 100 chronograms from the BEAST run used to generate the MCC tree in order to take into account phylogenetic uncertainty and differences in branch lengths between runs, with 10,000 simulations conducted on

each chronogram. The methods used to size-correct the variables and perform the PCA are dependent on the phylogeny, so for each tree these steps were repeated before the disparity through time analysis. These analyses were also repeated with size (standard length) as a variable (which had been logged and averaged).

5.3.4 Convergence analysis

In order to assess if different clades were evolving under the same selective regime to the same adaptive peak phenotypic convergence was assessed using the R package SURFACE (a recursive acronym for SURFACE Uses Regime Fitting with Akaike Information Criterion (AIC) to model Convergent Evolution) (Ingram and Mahler, 2013). This method uses the OU process and begins with a model where every clade is in a single selective regime, in which species are attracted to a single adaptive peak, before the stepwise addition of selective regime shifts until no further regime shift improves the fit of the model (using finite sample size corrected AIC (AICc)). Accepting all model updates that improve the AICc has been shown to generally provide good performance (Ingram and Mahler, 2013; Mahler et al., 2013). Following this stage the 'backwards' phase of the model evaluates if the AICc score can be improved further by allowing any of the selective regimes to be collapsed together meaning that they share the same adaptive peak (multiple regime collapses were allowed at each step). In the SURFACE analysis each trait is assumed to evolve independently. The SURFACE analysis was conducted for the first four PC axes and subsequently repeated using both the first four PC axes and standard length (which had been natural logged and averaged).

Some degree of convergence, as identified by SURFACE, is likely to occur in a radiation simply by chance. In order to account for this, simulated data was used to detect if and where the extent of convergence in the original analysis was greater than that seen by chance when regime shifts are taken into account in the model. Data was simulated under the final model from the 'forward' phase of the SURFACE analysis, which includes multiple regime shifts to different adaptive peaks (but no convergence) 99 times and the SURFACE model (forwards and backwards) was run on each of these datasets. In addition null datasets were simulated using the parameters from the initial model in SURFACE (a single optimum OU model) without any regime shifts.

There is some phylogenetic uncertainty in the *Synodontis* phylogeny at both the tips and deeper in the phylogeny. To assess whether phylogenetic uncertainty played a role in the estimate of the extent of convergence SURFACE was also run on the same 100 phylogenies as used in the estimate of the MDI as described above.

5.3.5 Geographic Information System (GIS)

In order to investigate geographic biases in the collection of museum specimens, the collection localities were plotted using a Geographic Information System (GIS) developed in ESRI ArcMap 10. *Synodontis* specimen data was taken from two online databases (Fish2Net – 399 specimens and GBIF – 2,302 specimens) with additional specimens (not included in these datasets) added from four museum catalogues: Royal Museum of Central Africa (2,302 specimens), The American Museum of Natural History (AMNH) (309 specimens), The Smithsonian Museum (57 specimens) and the South African Institute of Aquatic Biodiversity (SAIAB) (824 specimens). In total specimens from 47 different institutions were mapped. For 4,556 specimens, latitude and longitude data was included in the museum accession information; however in many cases only a description of the locality was available. In such cases a three-step approach was taken to assign geographic coordinates based on the locality text string. Firstly, locality text strings identical to one in the located dataset were assigned the same coordinates. For items without an exact match, string matching, both on single words and word pairs in the locality string was attempted against the located dataset to generate a 'shortlist' of potential locations, which were subsequently checked manually. In cases where no close match to an existing location was found, locality strings were geo-referenced manually using the GEOLocate Web Application and/or Google Maps. Localities that could not be identified were excluded.

Species names for each record were checked against the Catalog of Fishes (Eschmeyer, 2014) to identify and correct synonyms. Records not identified to species level, as well as those recorded as affiliated (aff.) or similar to (cf.) were excluded from the dataset. The resulting geolocated dataset consisted of 4,556 specimens with primary locations, 464 specimens located through string matching and 1,174 manually located specimens giving a total of 6,194 records, representing 126 different *Synodontis* species.

Bioclimatic and altitude data for Africa was taken from www.worldclim.org (Hijmans et al., 2005) with a resolution of 2.5 arc-minutes. This data is representative of conditions from 1950 to 2000. There are 19 bioclimatic indicators, shown in Table 1. Data from each of these raster layers was used to calculate the predicted climatic range of each of the *Synodontis* species found to be convergent in the SURFACE analysis using maximum entropy modelling of species geographic distributions (MaxEnt) (Phillips et al., 2006). The MaxEnt analysis was also carried out for all datapoints to represent the predicted climatic range of the genus as a whole. The distribution is based on the balance threshold computed by MaxEnt, this method reserves a randomly chosen 25% of the dataset which is tested against the climatic range predicted by the remainder of the dataset, in order to calculate the omission rate. The balance threshold is calculated based on this omission rate, the cumulative threshold and the fractional predicted area.

Table 1 Explanation of bioclimatic indicators

Indicator	Description
BIO1	Annual Mean Temperature
BIO2	Mean Diurnal Range (Mean of monthly (max temp - min temp))
BIO3	Isothermality (BIO2/BIO7) (* 100)
BIO4	Temperature Seasonality (standard deviation *100)
BIO5	Max Temperature of Warmest Month
BIO6	Min Temperature of Coldest Month
BIO7	Temperature Annual Range (BIO5-BIO6)
BIO8	Mean Temperature of Wettest Quarter
BIO9	Mean Temperature of Driest Quarter
BIO10	Mean Temperature of Warmest Quarter
BIO11	Mean Temperature of Coldest Quarter
BIO12	Annual Precipitation
BIO13	Precipitation of Wettest Month
BIO14	Precipitation of Driest Month
BIO15	Precipitation Seasonality (Coefficient of Variation)
BIO16	Precipitation of Wettest Quarter
BIO17	Precipitation of Driest Quarter
BIO18	Precipitation of Warmest Quarter
BIO19	Precipitation of Coldest Quarter

5.4 Results

5.4.1 Principal Component Analysis (PCA) loadings

The first six principal components explain 94.2% of the variance with the first four axes explaining 81.2% of the variance (42%, 15.2%, 13.7% and 10.3% respectively) in the phylogenetic principal component analysis (PCA). For further analyses only the first four principal components were retained, as PC5 (6.8% of the variance) and PC6 (6.2% of the variance) were difficult to interpret when the loadings are examined. High scores on PC1 correspond to shorter head lengths, smaller head heights, smaller distances from the base of the head to the centre of the eye, smaller snout length, smaller body depth and smaller body width and this axis corresponds broadly to the first PC axis from Chapter 4. On the second PC axis high scores correspond to longer snout widths, smaller caudal peduncles, longer head lengths, and longer snout lengths. The third PC is mainly influenced by snout length with higher scores representing smaller snout widths. The fourth PC is most heavily influenced by the caudal peduncle with higher scores corresponding to larger caudal peduncles.

PC1 of the PAST analysis explains 43.5% of the variance, loadings are generally lower than for the phylogenetic PCA but are broadly concordant with those results, with the loadings spread over multiple features, with the exception of snout width, which does not contribute to PC1 in the PAST analysis but does in the phylogenetic PCA. The loadings for PC2 (17.2% of the variance) are also similar between the two analyses with the exception of head length which contributes strongly to the axis in the phylogenetic analysis but not in the PAST analysis. Snout length is a heavy loading on PC3 in both analyses, however body width is important in the PAST analysis but not in the phylogenetic PCA. PC4 is most strongly influenced by the caudal peduncle width in both analyses.

5.4.2 Morphospace

In general the different areas of morphospace are occupied by taxa from multiple different geographic areas in Africa (Figure 1). The species from the Congo river system are the most widespread in morphospace while in East Africa there are multiple species that occupy very similar parts of morphospace. When the PC axes are considered separately this pattern holds, with the East African taxa tightly

clustered whereas the West African taxa and taxa from the Congo River basin are more widely distributed in morphospace. Taxa from West Africa found in the Lower Guinea Forest are also tightly clustered in a relatively small area of morphospace (Figure 1).

The taxa from LT are well distributed in morphospace and *S. njassae*, the only species sampled from Lake Malawi, occupies a similar area of morphospace to both taxa from riverine environments and some species from LT. In the morphospace described by PC1 and PC2 *S. njassae* occupies a very similar area of morphospace to two LT species (Clade 2, Clade 5) and is similar to *S. multipunctatus* on PC3 and multiple species on PC4. Some of the LT taxa have similar values on all retained PC axes and clustering of the LT taxa is most visible when PC1 and PC3 are plotted (Figure 1), however, on other axes they are more widely distributed and in the PC1 and PC2 morphospace occupy more area than all of the remaining East African taxa.

When the additional species not included in the phylogeny were included and the PCA performed on the entire dataset the additional species in this dataset were found to be very close in morphospace to the specimens from the phylogeny across PC axes (Figure 2), indicating no substantial increase in morphospace by including these taxa. This suggests that the results of the phylogenetic PCA are not biased to include only a subset of morphospace and the results of the phylogenetic PCA were used for further analyses.

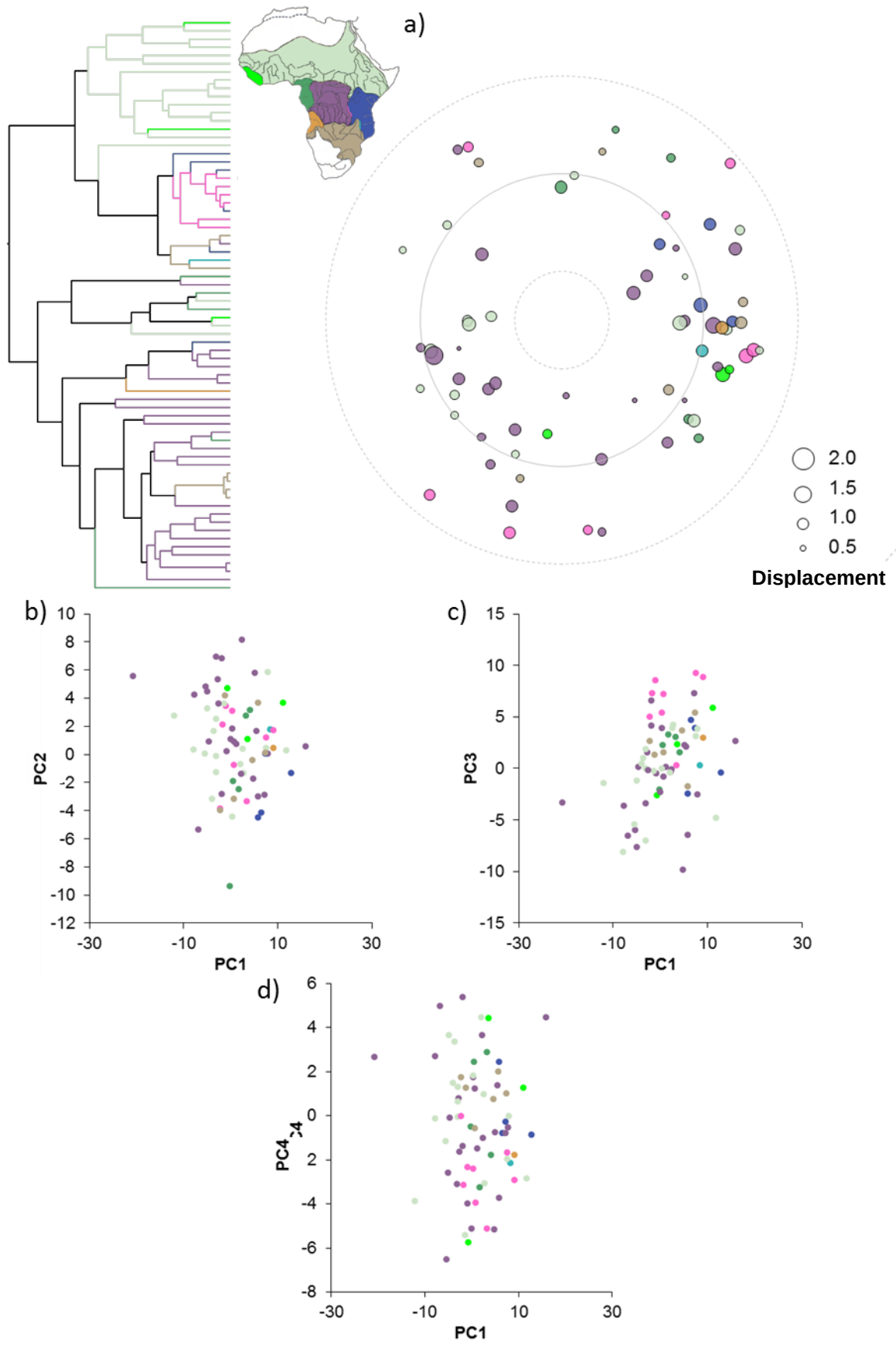


Figure 1 a) Morphospace wheel plot representing PC axes 1-3, b) plot of PC axes 1 and 2, c) plot of PC axes 1 and 3 and d) plot of PC axes 1 and 4. Colours represent the locality of the samples as shown on the map (Lake Tanganyika is pink whereas Lake Malawi is turquoise) and cladogram shows the phylogenetic placement of different collection localities.

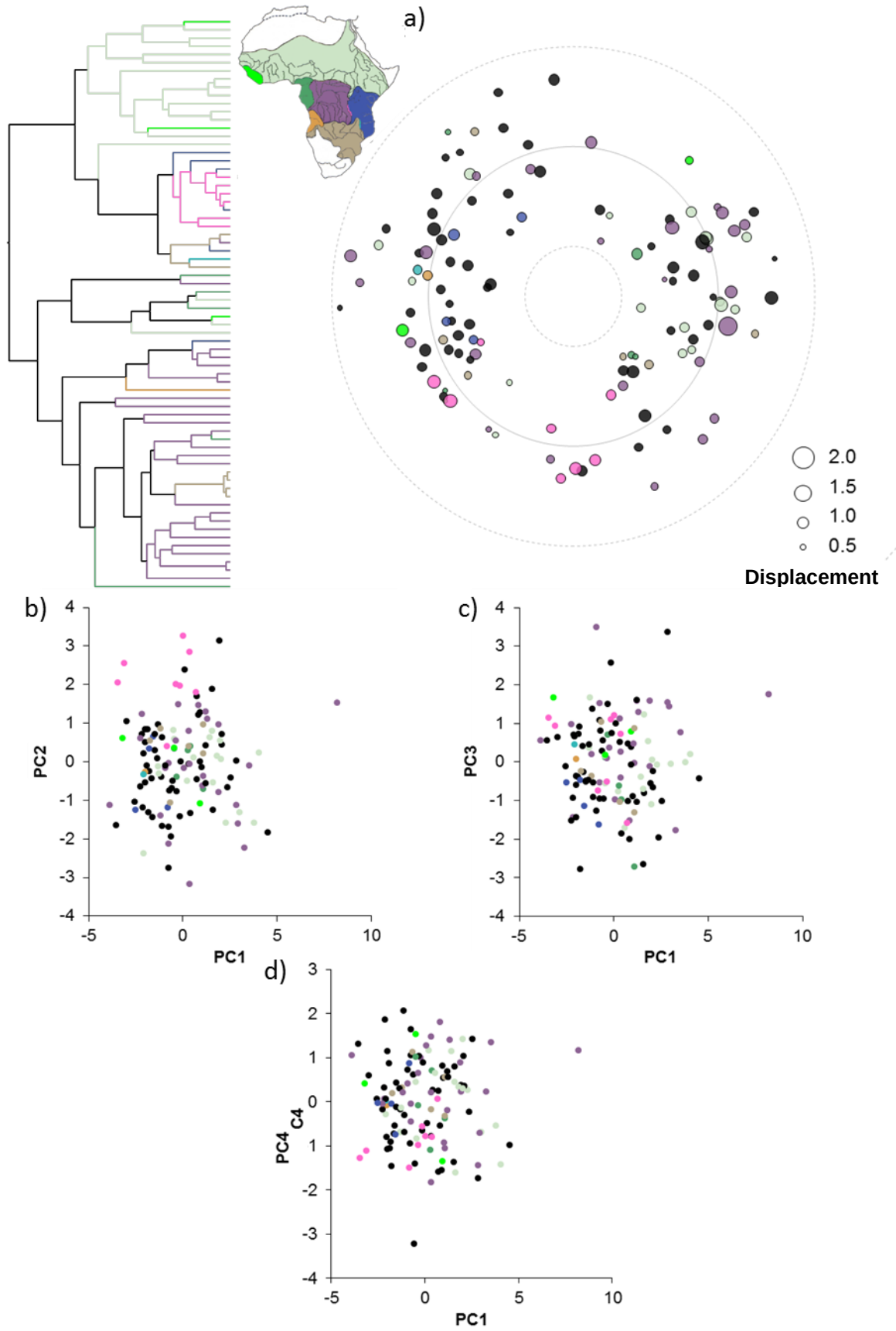


Figure 2 a) Morphospace wheel plot representing PC axes 1-3, b) plot of PC axes 1 and 2, c) plot of PC axes 1 and 3 and d) plot of PC axes 1 and 4. Colours represent the locality of the samples as shown on the map (Lake Tanganyika is pink whereas Lake Malawi is turquoise) and cladogram shows the phylogenetic placement of different collection localities. Samples not in the phylogeny are shown in black.

Table 2 Results of model fitting for PC axes 1-4 and standard length

	Brownian motion		Ornstein-Uhlenbeck		Early burst	
	dAICc	Akaike weight	dAICc	Akaike weight	dAICc	Akaike weight
PC1	11.452	0.003	0.000	0.996	13.640	0.001
PC2	21.397	0.000	0.000	1.000	23.584	0.000
PC3	6.867	0.031	0.000	0.959	9.054	0.010
PC4	20.567	0.000	0.000	1.000	22.755	0.000
Standard Length	30.407	0.000	0.000	1.000	32.595	0.000

5.4.3 Disparity through time

Akaike weights suggest that the OU model is preferred for each PC axis (Table 2), which is the starting model for the SURFACE analysis. There is higher subclade disparity than expected under Brownian motion throughout the entire radiation with much of it above the 95% confidence levels when considering PC1-4 together (Figure 3). This general pattern reflects that seen in PC1, which shows a peak around a quarter of the way through the radiation but is only above the 95% confidence interval in the last 10% of the radiation. The signal is much stronger in the PC2 plot where there is a clear peak a quarter of the way through the radiation which remains above the 95% confidence level from this point. For PC3 the signal is more similar to Brownian motion (mainly within the 95% confidence levels) but also shows greater subclade disparity than expected. The pattern in PC4 shows that there is greater subclade disparity than expected under Brownian motion throughout the radiation but 60% of the way through there is a large peak and the data is above the 95% confidence level from this point. The pattern in the standard length plot echoes that seen in PC4. In all the disparity through time plots (Figure 3) there is a deviation from Brownian motion towards the present, which may reflect incomplete taxon sampling rather than a genuine pattern. The MDI values are positive for all analyses with those for PC1-4 combined, PC2, PC4 and standard length significant at the 95% level (Table 3).

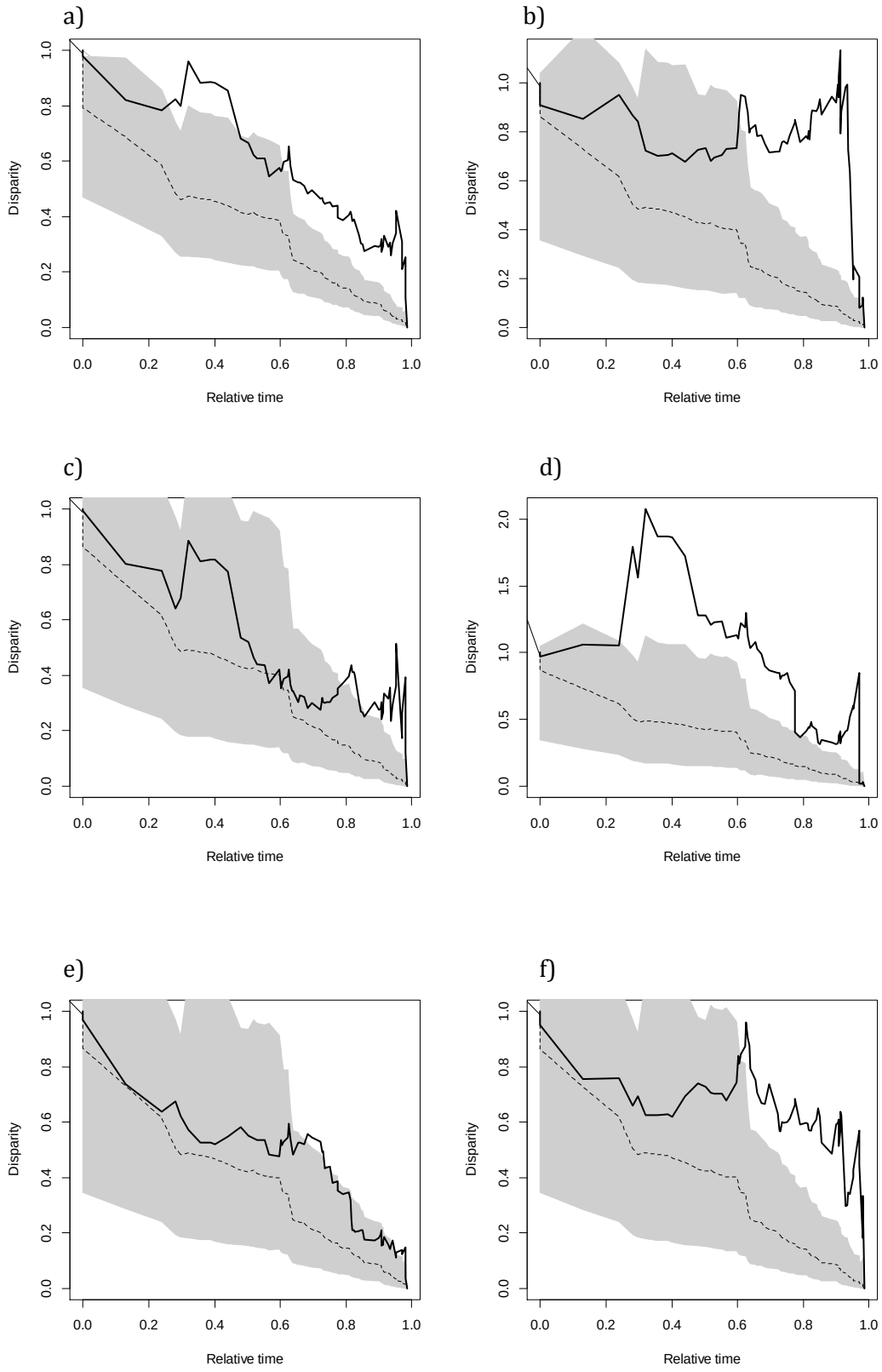


Figure 3 Disparity through time plots for a – PC1-4, b – Standard length, c – PC1, d – PC2, e – PC3 and f – PC4 based on the MCC tree. The grey area shows 95% confidence intervals, the black line, disparity, and the dashed line predicted disparity under Brownian motion.

Table 3 MDI values for the MCC tree and across a subsample of 100 trees

	MCC tree		100 trees			
	MDI	P-value	MDI mean	MDI standard deviation	P-value mean	P-value standard deviation
PC1	0.160	0.140	0.181	0.023	0.110	0.029
PC2	0.649	<0.001	0.533	0.189	0.012	0.030
PC3	0.115	0.220	0.191	0.142	0.173	0.153
PC4	0.269	0.033	0.274	0.049	0.036	0.026
PC1-4	0.254	0.001	0.268	0.022	0.001	0.001
Standard length	0.374	0.005	0.268	0.022	0.001	0.001

5.4.4 SURFACE analysis

In the SURFACE analysis to detect convergence in adaptive regimes within the radiation k=8 regime shifts to different adaptive peaks were found during the 'forward' phase. During the 'backward' phase these were collapsed to k'=6. The AICc improved from 1537.6 to 1495.4 during the 'forward' phase and then further improved to 1484.9 during the 'backward' phase. The number of shifts towards convergent regimes occupied by multiple lineages (c) was 3 meaning that the proportion of shifts that are convergent, measured as c/k, is 0.375 (Table 4). The phylogenetic half-lives for PC1=2.88 Ma, PC2=8.67 Ma, PC3=2.44 Ma and PC4=0.03 Ma, show large variation between PC axes and represent 8.3%, 24.99%, 7.03% and 0.09% of the tree respectively. The rates of stochastic trait evolution also vary between PC axes PC1 σ^2=12.3, PC2 σ^2=1.4, PC3 σ^2= 4.1, PC4 σ^2=199.1. The estimated phylogenetic half-life is very low for PC4 suggesting very rapid evolution to the selective optimum, however, the rate of stochastic evolution is estimated to be very high, reducing the distinctness of regimes.

When standard length is included in addition to PC axes 1-4 fewer regime shifts (k=6) were detected during the 'forward' phase and these were collapsed to k'=5 during the 'backward' phase. In this analysis the AICc improved from 1579.6 to 1531.8 during the 'forward' phase and further improved to 1525.3 during the 'backward' phase. The number of convergent shifts towards regimes occupied by multiple lineages is two which gives a similar proportion of convergent shifts (c/k=0.333) to the analysis without standard length (Table 4). In this analysis the phylogenetic half-lives, standard length=1.94 Ma, PC1=3.18Ma, PC2=11.2Ma, PC3=3.05Ma, PC4= 0.04Ma, show the same patterns as that seen in the analysis without standard length. The rates of stochastic trait evolution also show the same

pattern, standard length σ^2=0.06, PC1 σ^2=11.4, PC2 σ^2=1.4, PC3 σ^2= 3.6, PC4 σ^2=166.3.

There is one convergent regime that evolved multiple times in each analysis (Figure 4 and Figure 5). This regime applies to taxa found in both the Congo basin and the Nilo-Sudan region and only the presence/absence in this regime of *S. alberti* differs between the analysis with or without standard length. There are two additional rate shifts in the Congo basin (Figure 4 and Figure 5) that found are in both analyses, one containing *S. woosnami* and *S. macrostigma* and another containing *S. nigriventris* and *S. contractus*. In both analyses there is also a different selective optimum in the East African Clade. When just the PC axes are considered the entire clade is included in this rate shift whereas when body size in included only the unnamed LT taxa are included. The trait optima for the convergent regime are close to the trait values in both analyses (Figure 4 and Figure 5), however the selective optima for the regime shift leading to the species *S. macrostigma* and *S. woosnami* was far from the trait values for this species which is most visible in PC2. The extent of convergence was not exceptional in either analysis (Figure 6) with similar values for c in the analyses with both the model for the 'forward' SURFACE run and for the single optima starting OU model.

Table 4 Calculated parameters from the SURFACE analysis for PC axes 1-4 both with and without standard length

	PC axes 1-4 only		PC axes 1-4 and standard length	
	MCC tree	**100 trees**	**MCC tree**	**100 trees**
k	8	8.80 (±2.06)	6	6.28 (±1.50)
k'	6	6.14 (±1.26)	5	5.02 (±0.85)
c	3	4.89 (±1.92)	2	2.41 (±1.72)
c/k	0.38	0.54 (±0.12)	0.33	0.34 (±0.22)

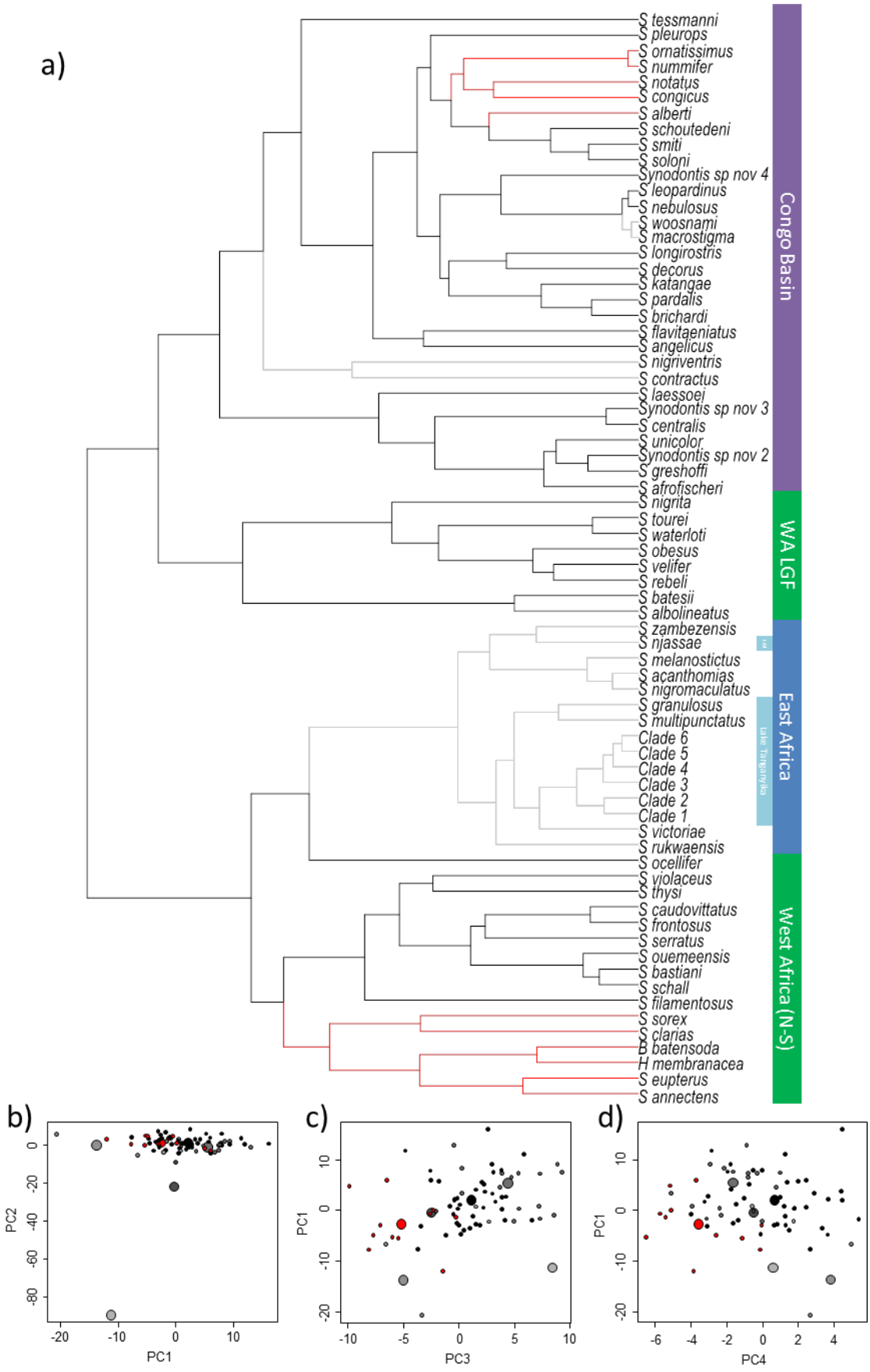


Figure 4 Results of the SURFACE analysis for PC1-PC4, a) Cladogram with regime shifts in grey and convergent regime shifts coloured red. Figure panels show how trait values (small circles) cluster around trait optima (large circles) for b) PC1 and PC2, c) PC3 and PC1 and d) PC4 and PC1.

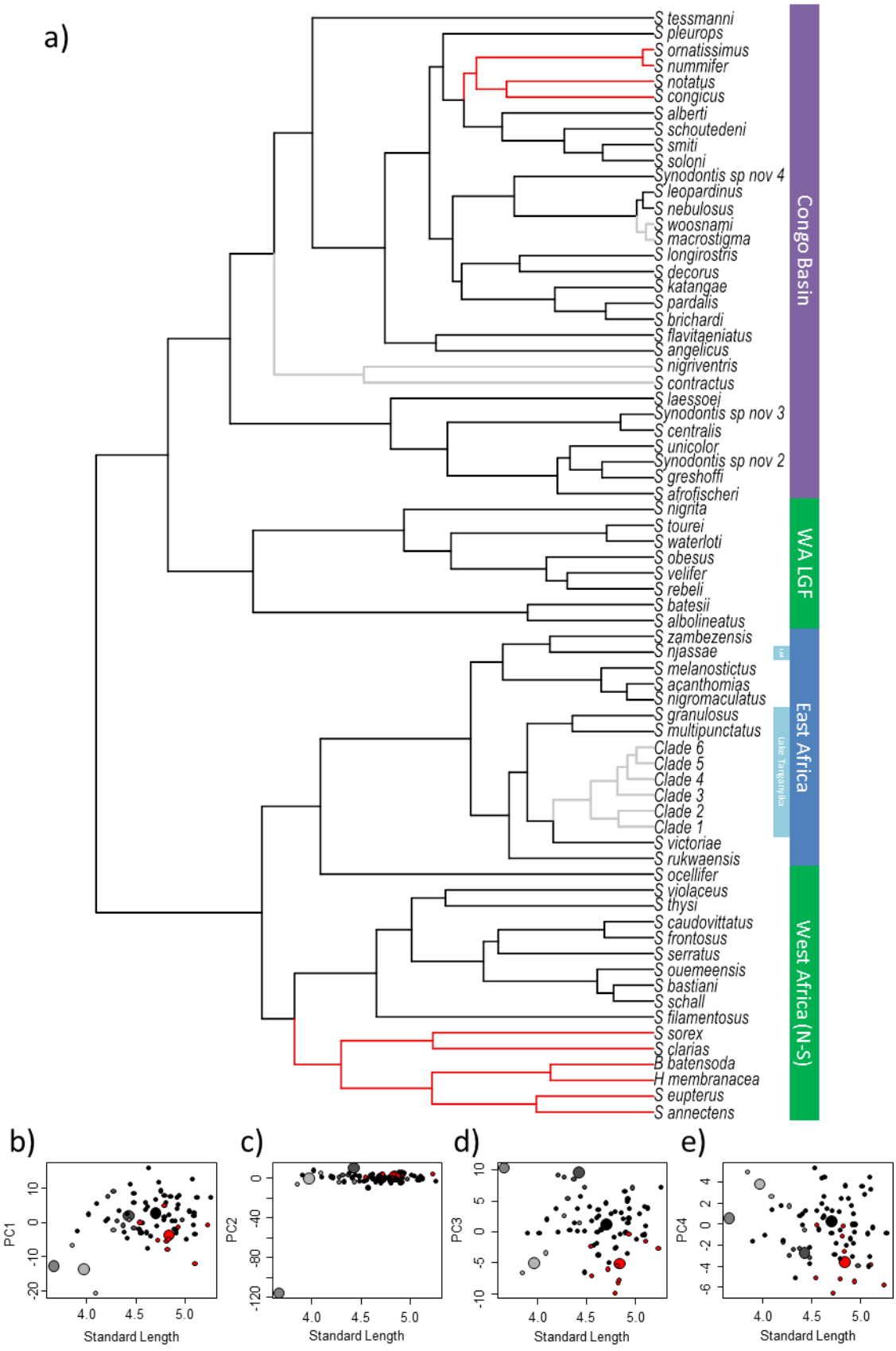


Figure 5 Results of the SURFACE analysis for standard length and PC1-PC4 a) Cladogram with regime shifts in grey and convergent regime shifts coloured red. Figure panels show how trait values (small circles) cluster around trait optima (large circles) for b) standard length and PC1, c) standard length and PC2, d) standard length and PC4 and e) standard length and PC4.

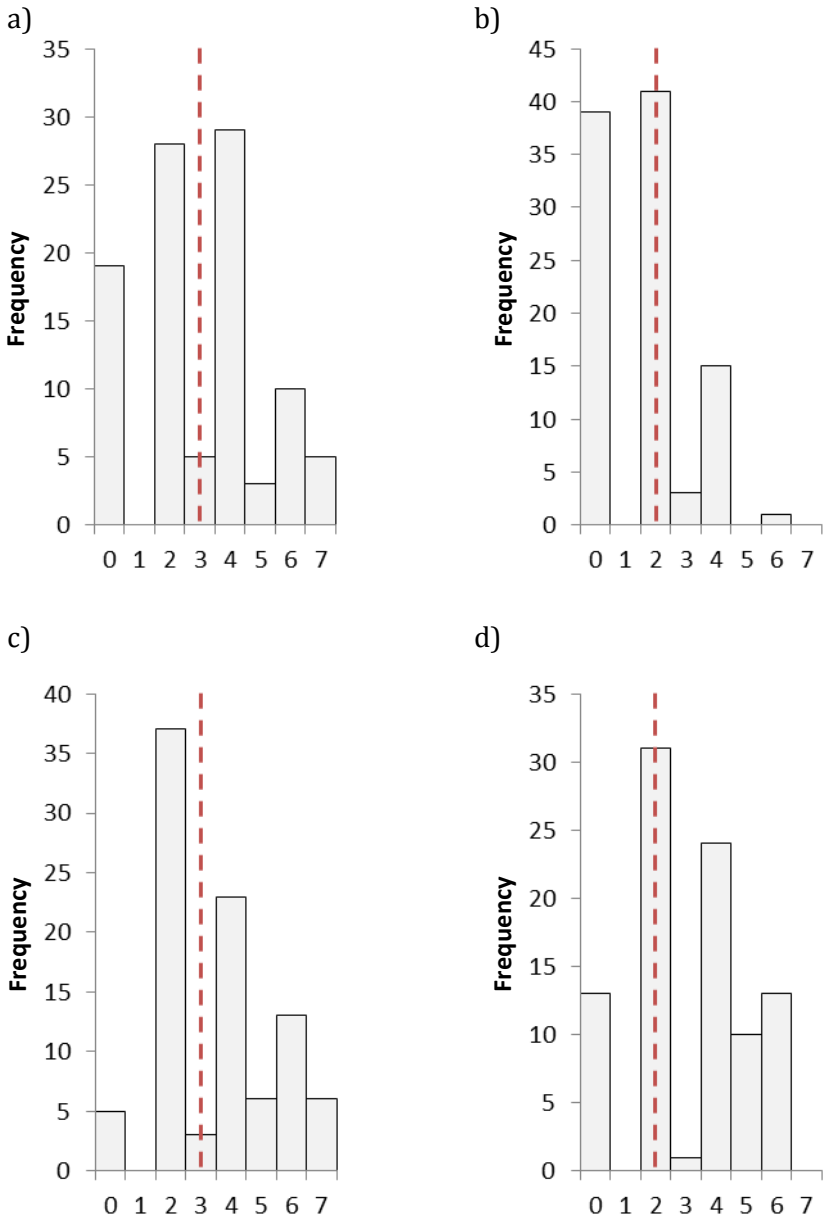


Figure 6 Histograms showing values of c (the number of shifts towards convergent regimes occupied by multiple lineages) from 99 simulated datasets with the red line indicating the value of c calculated for the actual dataset. Charts a and b show the results using the final model from the 'forward' phase of the SURFACE analysis for PC1-4 (a) and PC1-4 plus standard length (b). Charts c and d show the results using a model with a single adaptive peak (the initial OU model in SURFACE) for PC1-4 (c) and PC1-4 plus standard length (d).

5.4.5 Geographic distribution of museum collections

Occurrence data for the eleven species found to be convergent in the SURFACE analysis (Figure 4), with *S. alberti* coded as 'Clade A' in this map, shows large ranges for both clades and considerable but not total range overlap (Figure 7). The estimated climatic range is also very similar between these clades with many of the species predicted to occur in the same locations, as seen by the extent of the red area in which all convergent species are predicted based on their climatic distributions (Figure 7).

Occurrence data for all specimens shows that although *Synodontis* specimens are held in institutions around the world (Figure 8) data from multiple institutions are needed in order to conduct a study at the continental scale as the collections of individual institutions are clustered geographically. There is, however, sampling from throughout the predicted geographic range of *Synodontis* with Angola, northern Mozambique, parts of the species rich Congo basin and South Sudan identified as potentially under sampled regions. Morphological data could not be collected for 18 of the species in the mapped dataset. The distributions of these species is broadly congruent with the remainder of the genus (Figure 8), suggesting that the absence of these species does not lead to a geographical bias in the morphological dataset.

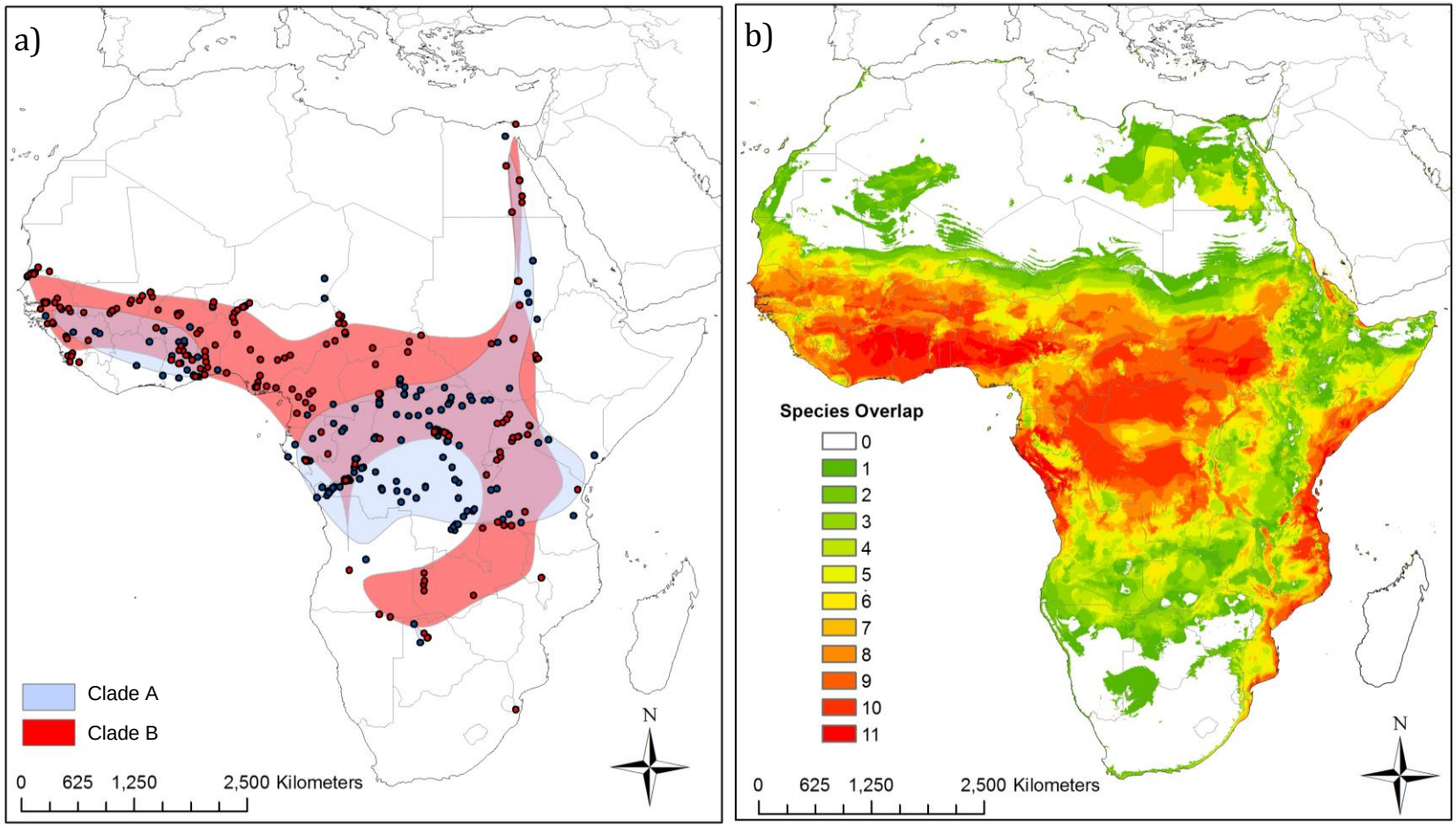


Figure 7 a) Sampled range of the convergent taxa. The group called 'Clade A' includes the topmost convergent (red) clade in Figure 4, plus the closely related *S. alberti*. 'Clade B' is the bottommost convergent clade in Figure 4. Shaded areas are polygons created by aggregating points at a distance of 1,000km. b) Overlap of predicted climatic distribution of all 11 species in clades A and B (MaxEnt, balance threshold).

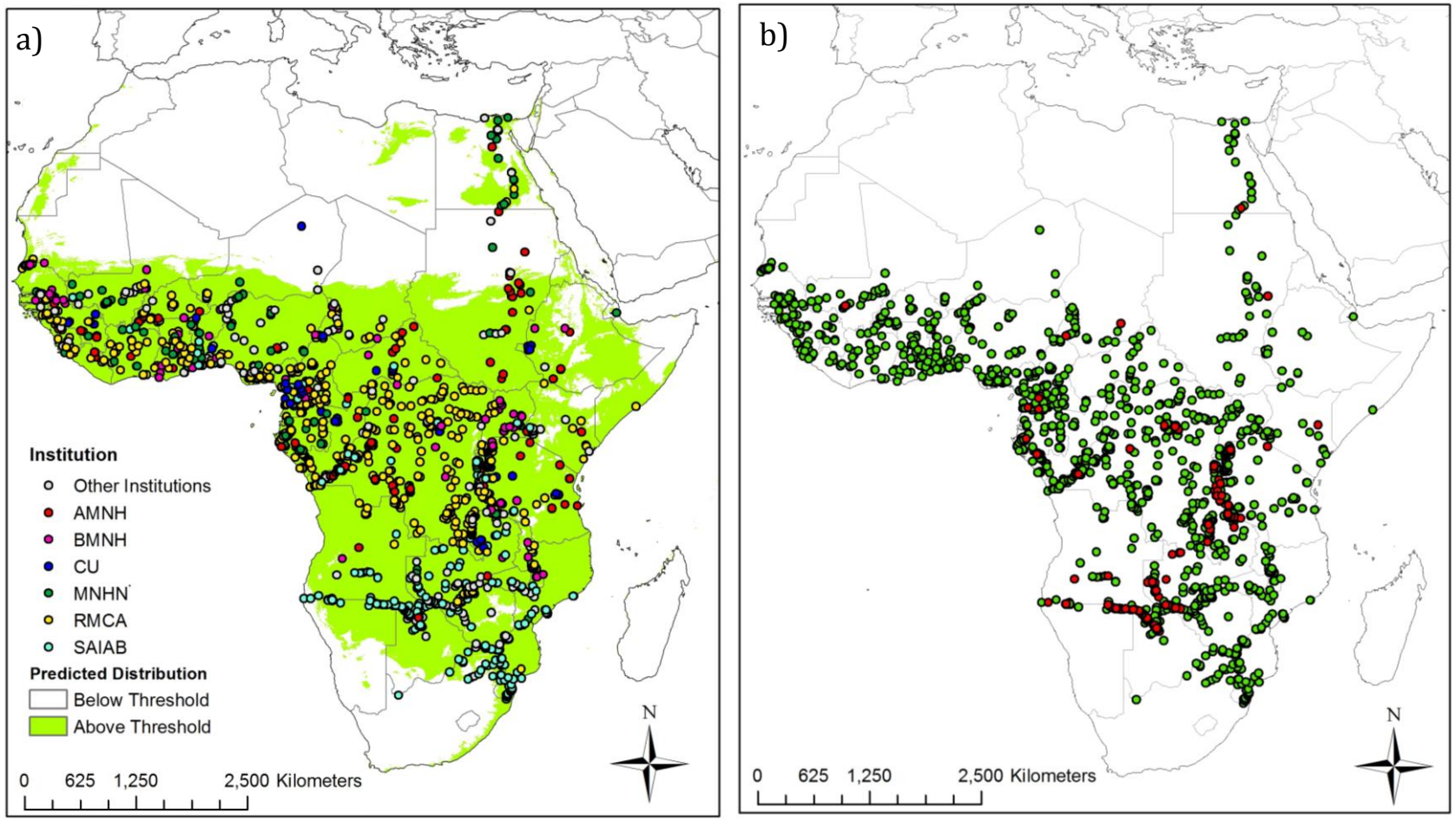


Figure 8 a) Predicted climatic range of *Synodontis* (MaxEnt balance threshold) vs actual sampled data. Points are coloured by institution (AMNH – American Museum of Natural History, BMNH – Natural History Museum, London, CU – Cornell University Museum of Vertebrates, MNHN - Muséum National d'Histoire Naturelle, Paris, RMCA – Royal Museum of Central Africa, SAIAB – South African Institute for Aquatic Biodiversity. b) Specimen data coloured by presence/absence of morphological data for that species in this study. Green – presence, red – absence.

5.5 Discussion

There is considerable overlap in morphospace of species from different geographic areas and increased diversity within subclades suggesting that the independent evolution of taxa in different drainage basins produced similar morphologies. However, there is little evidence for morphological convergence suggesting that phenotypic constraints are responsible for this pattern.

5.5.1 Lake Tanganyika taxa

There is no evidence that LT taxa have distinct or extreme morphologies compared to the riverine taxa. In the non-phylogenetically corrected PCA plots (Figure 2), the LT taxa cluster together, but fall within the broader spread of species. While this analysis suggests that without accounting for their recent origin they are relatively similar to each other, they are not morphologically distinct from the riverine *Synodontis*. When phylogeny is taken into account the LT taxa are widespread in morphospace (Figure 1), showing a far greater spread than the remainder of the East African taxa within which they are nested. While they do show divergent phenotypes on PC3, in general their morphologies cluster with the riverine taxa suggesting a recent and relatively rapid, increase in morphological disparity in LT but toward forms similar to those found in rivers. This is consistent with their habitats, as riverine *Synodontis* tend to be found in mature rivers, which may be capable of supporting phenotypes also viable in lacustrine conditions and Lake Albert is known to contain riverine species. It is also supported by the findings of Chapter 4 that within the LT radiation there is more disparity within subclades than expected by Brownian motion suggesting that phenotypes within LT are evolving rapidly compared to those in riverine environments.

5.5.2 Morphological change through time

The disparity through time analysis shows positive MDI values for all axes with PC2, PC4, standard length and combined PC1-4 all significantly higher than the expectation under Brownian motion, with results robust to uncertainty in branch lengths and topology. These results indicate that a large proportion of the total disparity is partitioned within subclades, and as a result that subclades are more

likely to overlap in morphospace. This is consistent with a pattern of convergent evolution as taxa independently evolve to occupy similar areas of morphospace. These conclusions are not dependent on events late in the radiation, suggesting that the observed pattern is resilient to the potential effects of incomplete taxon sampling, which can lead to an increase in the perceived disparity of younger nodes.

These results concur with those of (Harmon et al., 2003) who found that Iguanian lizards taxa with steady rates of lineage accumulation, as seen in *Synodontis*, exhibit high within subclade variation and predicted this to be a general pattern. In these Iguanian lizards this pattern was found in the South American genus *Liolaemus*, in which subclades rarely overlap geographically allowing each to diversify independently, the same pattern as observed in *Synodontis*. There are few studies of this at large spatial scales and the results vary, for example Old World lacertid lizards show the same pattern as *Synodontis* with constant rates of lineage accumulation accompanied by disparity distributed within subclades (Hipsley et al., 2014) whereas Neotropical birds in the family Furnariidae show constant rates of lineage accumulation but morphological disparity is mostly distributed between subclades (Derryberry et al., 2011).

In a classic model of adaptive radiation there are several stages of diversification, for example habitat segregation followed by trophic partitioning then perhaps further diversification driven by other factors such as sexual selection (discussed in Salzburger, 2009). This can also lead to the pattern of increased disparity within subclades if this disparity is related to the second stage of diversification. This pattern is seen in the Madagascan passerine bird radiation of vangid species, where species initially diverged in size (disparity partitioned between clades) and then subsequently diversified in beak morphology leading to species from different subclades occupying similar areas of beak morphospace (Jønsson et al., 2012). This shows the importance of identifying traits of adaptive divergence, including multiple traits from throughout the body, otherwise adaptive divergence can be missed and only subsequent convergence identified. There is however no evidence of this in the *Synodontis* as traits were taken from throughout the body and PC axes represent changes in multiple different traits. Further support for the pattern identified in *Synodontis* is provided by the results for standard length which also show increased disparity within subclades. This is

despite the potential for biases towards smaller specimens being accessioned in museum collections due to the travel and storage difficulties associated with intact large specimens, and the paucity of size information for *Synodontis* in the literature.

5.5.3 Evidence for convergent evolution

The disparity through time plots indicate that there is high disparity within subclades which may cause subclades to overlap with each other in morphospace. This is corroborated by the PCA plots which show that taxa from multiple geographic areas occupy similar areas of morphospace with only the East African taxa showing any evidence of morphological clustering. These plots show a broad range within morphospace rather than distinct clusters suggesting that there is broad variation in the clade rather than the distinct phenotypes often seen in adaptive radiations and associated with different niches (e.g., Losos et al., 1998; Muschick et al., 2012). When the presence of convergence is explicitly investigated using SURFACE (Ingram and Mahler, 2013) there is little convergence of adaptive optima between geographic areas with only one convergent regime identified, the occurrence of which can be explained by chance alone. This is despite the finding that SURFACE tends to slightly overestimate the number of regimes and shifts to those regimes when there were few regime shifts (Grundler and Rabosky, 2014; Ingram and Mahler, 2013). Missing data is a potential problem in the SURFACE analysis and can lead to an underestimate of rate shifts (Ingram and Mahler, 2013), however when *Anolis* lizards were studied parameter estimates were similar in runs using 60% of the original dataset (Mahler et al., 2013), a level that corresponds to the degree of taxon sampling in this study. It is therefore not clear whether the low numbers of regime shifts identified in this study are influenced by missing data and further data to extend the phylogeny would increase confidence in the results. The *Synodontis* clades that are identified as convergent by SURFACE do share some characteristics, both are wide-ranging, overlap geographically over a large part of their range and are predicted to have similar potential geographic distributions based on climate and altitude data (Figure 7).

Where convergence has been detected in other studies using SURFACE it coincides with distinct changes in ecology, for example, the evolution of ecotypes in *Anolis* lizards (Mahler et al., 2013) or shifts unrelated to diet such as habitat use

or locomotion in snakes from North America and Australia (Grundler and Rabosky, 2014). In-depth diet information for *Synodontis* is largely lacking but many species are found in similar environments in mature rivers and the broad spread of taxa in morphospace does not suggest step changes in ecology. The lack of convergence and lack of distinct PCA clusters in *Synodontis* may be explained by the presence of strong constraints on morphological evolution throughout the genus. *Synodontis* species have a conserved body plan and are often difficult to identify morphologically, as discussed in Chapter 4 for the LT taxa. The pattern of morphological change was best explained by the application of an OU model (or multiple OU models in the SURFACE analysis), a common model for comparative data that is associated with constraints shaping evolution (Harmon et al., 2010).

5.5.4 Colonisation of Southern Africa

The expansion of taxa into new environments, such as different river basins, may play a role in driving morphological adaptation through a change in the selective regime. This is evidenced by the polyphyletic taxa that have recently colonised the Zambesi ichthyoprovince and thus southern Africa. In this phylogeny there are six taxa classed as belonging to the Zambesi ichthyoprovince in the analysis of Day et al., (2013) and four of these taxa are identified as belonging to selective regimes different to the majority of taxa, *S. nigromaculatus* and *S. zambezensis* are associated with the rate shift in the East African clade (not found when size is included) whereas *S. macrostigma* and *S. woosnami* are involved in an additional rate shift but their trait values are not close to the selective optima. This may indicate on-going diversification but this requires further research. It has been hypothesised that high morphological diversity in haplochromine cichlids in the rivers of Southern Africa arose in the now extinct Lake Makgadiikgadi (Joyce et al., 2005) but there is no evidence for this in *Synodontis* given the polyphyletic nature of the assemblage and the similar morphologies seen between taxa from Southern Africa and distant river basins.

5.5.5 Geographical sampling of Synodontis specimens

The additional species that were not in the molecular phylogeny were sampled from the same geographical areas as those included in the molecular phylogeny suggesting no bias between the molecular and specimen museum collections

despite the molecular collections generally being collected much more recently. Geographic sampling of *Synodontis* shows broadly good coverage of the predicted extent of the genus (based on climatic and altitude data) suggesting that the results in this study are not biased by excluding species from a single geographic region, with specimens broadly distributed between institutions in Africa, Europe and the USA. However, Day et al., (2013) found large genetic divergences within some taxa with large ranges and it is clear that individual institutions have geographically restricted collections which may be a limitation to large geographic studies of morphology. For example, this study contains measurements from specimens in five different institutions but may be missing morphological variation towards the southern part of the ranges of widespread species because the majority of specimens from Southern Africa are in the South African Institute for Aquatic Biology and it was not possible to measure these specimens for this study. The maps of collection data also show some specific areas for which collection data is sparse including Angola, particularly in the north, parts of the Congo basin, South Sudan and northern Mozambique which may act as a guide for potential collection trips to target under-sampled regions. A broad analysis of the sampling localities of museum collections for all African freshwater fish would suggest geographic areas that have been historically under-sampled in total as this analysis may be influenced by regions in which *Synodontis* show low abundance/diversity despite being predicted to occur there by climatic models.

5.5.6 Conclusions

Studying the LT *Synodontis* radiation in the context of their riverine relatives suggest that although morphological diversification has been rapid considering their recent origin, no distinct morphologies have evolved in lacustrine settings. Indeed, these lacustrine taxa have very similar morphologies to riverine taxa. A change in selective regime following the recent colonisation of a new environment appears to be a more general pattern in *Synodontis* and is also observed in the recent colonisation of rivers in Southern Africa. There is, however, little evidence of explicit convergence between taxa despite colonising very similar environments and this, coupled with their continuous distribution in morphospace, suggests that phenotypic constraints may be the dominant factor affecting morphological diversification in this group. The problem of incomplete sampling has hindered

wide scale studies of the patterns and processes of diversification across taxa at large spatial scales. In *Synodontis*, though the study of available records from museum collections does identify priority areas for future collecting, records suggest that the genus has been sampled from throughout its range. This broad sampling, including radiations in riverine systems and an ancient lake, suggests that *Synodontis* presents a useful system in which to study the patterns and processes of diversification in different environments.

5.6 References

Campagna, L., Benites, P., Lougheed, S. C., Lijtmaer, D. A, Di Giacomo, A. S., Eaton, M. D., & Tubaro, P. L. (2012) Rapid phenotypic evolution during incipient speciation in a continental avian radiation. *Proceedings of the Royal Society B: Biological Sciences*, **279**(1734), 1847–1856

Cooper, W. J., Parsons, K., McIntyre, A., Kern, B., McGee-Moore, A., & Albertson, R. C. (2010) Bentho-pelagic divergence of cichlid feeding architecture was prodigious and consistent during multiple adaptive radiations within African rift-lakes. *PloS One*, **5**(3), e9551

Danley, P. D., & Kocher, T. D. (2001) Speciation in rapidly diverging systems: lessons from Lake Malawi. *Molecular Ecology*, **10**(5), 1075–1086

Day, J. J., Bills, R., & Friel, J. P. (2009) Lacustrine radiations in African Synodontis catfish. *Journal of Evolutionary Biology*, **22**(4), 805–817

Day, J. J., Peart, C. R., Brown, K. J., Friel, J. P., Bills, R., & Moritz, T. (2013) Continental diversification of an African catfish radiation (Mochokidae: Synodontis). *Systematic Biology*, **62**(3), 351-365

Day, J. J., & Wilkinson, M. (2006) On the origin of the Synodontis catfish species flock from Lake Tanganyika. *Biology Letters*, **2**(4), 548–552

Derryberry, E. P., Claramunt, S., Derryberry, G., Chesser, R. T., Cracraft, J., Aleixo, A., & Brumfield, R. T. (2011) Lineage diversification and morphological evolution in a large-scale continental radiation: the neotropical ovenbirds and woodcreepers (aves: Furnariidae) *Evolution,* **65**(10), 2973–2986.

Eschmeyer, W. N. (Ed.) (2014) *Catalog of Fishes: Genera, Species, References.* <http://research.calacademy.org/research/ichthyology/catalog/fishcatmain.asp> Accessed on 19th August 2014

Fryer, G., & Iles, T. D. (1972) *The cichlid fishes of the Great Lakes of Africa*. Edinburgh: Oliver and Boyd

GEOLocate Web Application <http://www.museum.tulane.edu/geolocate/web/WebGeoref.aspx>

Google Maps <https://www.google.co.uk/maps/>

Grundler, M., & Rabosky, D. (2014) Trophic divergence despite morphological convergence in a continental radiation of snakes. *Proceedings of the Royal Society B: Biological Sciences*, **281**(1787) 20140413

Hammer, O., Harper, D. A. T., & Ryan, P. D. (2001) PAST: Paleontological Statistics software package for education and data analysis. *Palaeontologica Electronica*, **4**(1), 9

Harmon, L. J., Losos, J. B., Jonathan Davies, T., Gillespie, R. G., Gittleman, J. L., Bryan Jennings, W., & Mooers, A. Ø. (2010) Early bursts of body size and shape evolution are rare in comparative data. *Evolution*, **64**(8), 2385–2396

Harmon, L. J., Schulte, J. a, Larson, A., & Losos, J. B. (2003) Tempo and mode of evolutionary radiation in iguanian lizards. *Science,* **301**(5635), 961–964

Harmon, L. J., Weir, J. T., Brock, C. D., Glor, R. E., & Challenger, W. (2008) GEIGER: investigating evolutionary radiations. *Bioinformatics*, **24**(1), 129–131

Hijmans, R. J., Cameron, S. E., Parra, J. L., Jones, P. G., & Jarvis, A. (2005) Very high resolution interpolated climate surfaces for global land areas. *International Journal of Climatology*, **25**(15), 1965–1978

Hipsley, C., Miles, D., & Müller, J. (2014) Morphological disparity opposes latitudinal diversity gradient in lacertid lizards. *Biology Letters*, **10**, 20140101

Ingram, T., & Mahler, D. L. (2013) SURFACE: detecting convergent evolution from comparative data by fitting Ornstein-Uhlenbeck models with stepwise Akaike Information Criterion. *Methods in Ecology and Evolution*, **4**(5), 416–425

Irschick, D., Vitt, L., Zani, P., & Losos, J. (1997) A comparison of evolutionary radiations in mainland and caribbean Anolis lizards. *Ecology*, **78**(7), 2191–2203

Jønsson, K. A, Fabre, P.-H., Fritz, S. A, Etienne, R. S., Ricklefs, R. E., Jørgensen, T. B., & Irestedt, M. (2012) Ecological and evolutionary determinants for the adaptive radiation of the Madagascan vangas. *Proceedings of the National Academy of Sciences*, **109**(17), 6620–6625

Joyce, D., Lunt, D., Bills, R., Turner, G., Katongo, C., Duftner, N., & Seehausen, O. (2005) An extant cichlid fish radiation emerged in an extinct Pleistocene lake. *Nature*, **435***,* 90-95

Koblmüller, S., Sturmbauer, C., Verheyen, E., Meyer, A., & Salzburger, W. (2006) Mitochondrial phylogeny and phylogeography of East African squeaker catfishes (Siluriformes: Synodontis). *BMC Evolutionary Biology*, **6**, 49

Kozak, K. H., Weisrock, D. W., & Larson, A. (2006) Rapid lineage accumulation in a non-adaptive radiation: phylogenetic analysis of diversification rates in eastern North American woodland salamanders (Plethodontidae: Plethodon). *Proceedings of the Royal Society B: Biological Sciences*, **273**(1586), 539–546

Losos, J. B., Jackman, T. R., Larson, A., de Queiroz, K., & Rodriguez-Schettino, L. (1998) Contingency and determinism in replicated adaptive radiations of island lizards. *Science*, **279**(5359), 2115–2118

Lovette, I. J., Bermingham, E., & Ricklefs, R. E. (2002) Clade-specific morphological diversification and adaptive radiation in Hawaiian songbirds. *Proceedings of the Royal Society B: Biological Sciences*, **269**(1486), 37–42

Mahler, D. L., Ingram, T., Revell, L. J., & Losos, J. B. (2013) Exceptional convergence on the macroevolutionary landscape in island lizard radiations. *Science,* **341**(6143), 292–295

Montoya-Burgos, J. I. (2003) Historical biogeography of the catfish genus Hypostomus (Siluriformes: Loricariidae), with implications on the diversification of Neotropical ichthyofauna. *Molecular Ecology*, **12**(7), 1855–1867

Muschick, M., Indermaur, A., & Salzburger, W. (2012) Convergent evolution within an adaptive radiation of cichlid fishes. *Current Biology*, **22**(24), 2362–2368

Otero, O., & Gayet, M. (2001) Palaeoichthyofaunas from the Lower Oligocene and Miocene of the Arabian Plate: palaeoecological and palaeobiogeographical implications. *Palaeogeography, Palaeoclimatology, Palaeoecology*, **165**, 141–169

Partridge, T. C., Wood, B. A., & de Menocal, P. B. (1995) The influence of global climate change and regional uplift on large-mammalian evolution in East and Southern Africa. In E. S. Vrba, G. H. Denton, T. C. Partridge, & L. H. Burckle (Eds.), *Paleoclimate and evolution with emphasis on human origins.* New Haven: Yale University Press. pp. 331–355

Phillips, S. J., Anderson, R. P., & Schapire, R. E. (2006) Maximum entropy modeling of species geographic distributions. *Ecological Modelling*, **190**(3-4), 231–259

Pigot, A. L., Phillimore, A. B., Owens, I. P. F., & Orme, C. D. L. (2010) The shape and temporal dynamics of phylogenetic trees arising from geographic speciation. *Systematic Biology*, **59**(6), 660–673

Poll, M. (1971) Révision des Synodontis Africains (Famille Mochocidae) *Musee Royal de L'Afrique Centrale, Tervuren, Belgium, Annals Zoologiques*, **191**

Polly, P. D., Lawing, A. M., Fabre, A.-C., & Goswami, A. (2013) Phylogenetic principal components analysis and geometric morphometrics. *Hystrix*, **24**(1), 33–41

Rabosky, D. L., Santini, F., Eastman, J., Smith, S. A, Sidlauskas, B., Chang, J., & Alfaro, M. E. (2013) Rates of speciation and morphological evolution are correlated across the largest vertebrate radiation. *Nature Communications*, **4**, 1958

Revell, L. J. (2009) Size-correction and principal components for interspecific comparative studies. *Evolution*, **63**(12), 3258–3268

Ricklefs, R. E. (2012) Species richness and morphological diversity of passerine birds. *Proceedings of the National Academy of Sciences*, **109**(36), 14482–14487

Roger, J., Sen, S., Thomas, H., Cavelier, C., & Al Sulaimani, Z. (1993) Stratigraphic, palaeomagnetic and palaeoenvironmental study of the Early Oligocene vertebrate locality of Taqah (Dhofar, Sultanate of Oman) *Newsletters on Stratigraphy*, *28*(2-3), 93–119

Salzburger, W. (2009) The interaction of sexually and naturally selected traits in the adaptive radiations of cichlid fishes. *Molecular Ecology*, **18**(2), 169–185

Slater, G. J., Price, S. a, Santini, F., & Alfaro, M. E. (2010) Diversity versus disparity and the radiation of modern cetaceans. *Proceedings of the Royal Society B: Biological Sciences,* **277**(1697), 3097–3104

Vigliotta, T. (2008) A phylogenetic study of the African catfish family Mochokidae (Osteichthyes, Ostariophysi, Siluriformes), with a key to genera. *Proceedings of the Academy of Natural Sciences of Philadelphia,* **157**(1), 73–136.

Vreven, E., & Zamba, I. (2010) Synodontis carineae , a new species of mochokid catfish from the Kouilou-Niari River basin, Africa (Siluriformes: Mochokidae). *Ichthyological Exploration of Freshwaters*, **21**(4), 359–367

Wagner, C. E., Harmon, L. J., & Seehausen, O. (2012) Ecological opportunity and sexual selection together predict adaptive radiation. *Nature*, **487**(7407), 366–369

Wright, J. J. (2011) Conservative coevolution of Müllerian mimicry in a group of rift lake catfish. *Evolution*, **65**(2), 395–407

Zachos, J., Pagani, M., Sloan, L., Thomas, E., & Billups, K. (2001) Trends, rhythms, and aberrations in global climate 65 Ma to present. *Science,* **292**(5517), 686–693

Chapter 6
Conclusions

6.1 Summary of main results

This thesis provides the first study of the patterns and processes of diversification in the claroteine catfish radiation in Lake Tanganyika (LT), allowing this system to be compared and contrasted with both the *Synodontis* catfish radiation and other endemic radiations from this aquatic biodiversity hotspot. Through combined and parallel quantitative analyses of the claroteine and *Synodontis* radiations, an insight into different factors influencing catfish diversification in LT has been gained by placing them in the context of adaptive radiation, environmental change and present day processes. In the case of *Synodontis* the study of diversification in LT was extended to the continental scale allowing the similarities and differences between lacustrine and riverine morphological diversification to be investigated, as new environments offer different niches but similar morphological and genetic constraints exist.

The first molecular phylogeny of the Claroteinae in LT is presented in Chapter 2, based on a novel multigene dataset, enabling both the study of the diversification of these species in LT, and their use as a comparative system to other LT radiations, primarily *Synodontis,* as well as the well-studied cichlid fishes. By drawing upon this phylogeny, the factors involved in claroteine diversification have been investigated, with both evidence of adaptive divergence (Chapter 4) and strong evidence of recent isolation by distance within *Lophiobagrus cyclurus* (Chapter 3) being identified. While the evolutionary relationships within the sympatric *Synodontis* radiation had been well studied both within LT (Day and Wilkinson, 2006; Koblmüller et al., 2006) and at a larger spatial scale (Day et al., 2009; Day et al., 2013), this thesis represents the first investigation into the role of adaptive divergence within LT and of population genetics in the species *Synodontis multipunctatus*. In contrast to the claroteines, there is little evidence for adaptive divergence between clades with some convergence between them (Chapter 4) and genetic isolation by distance is not strong in the deep water *S. multipunctatus*. This pattern of morphological conservatism in the LT *Synodontis* is also seen at a continental scale (Chapter 5), where despite evidence of recent genetic

diversification within LT, the LT *Synodontis* display similar morphological features to their riverine congeners from throughout Africa.

There remains some taxonomic uncertainty within the LT catfish radiations, with morphological keys not accurately describing the different clades in the molecular phylogeny (discussed in Chapter 4), and parts of LT being under sampled for these species, making inferences into their modes of diversification problematic. However, the variable patterns of diversification observed in the LT catfish radiations reflect the complex patterns of diversification seen in LT cichlids, suggesting that a complex interplay of idiosyncratic factors are likely to be responsible for the range of present diversity.

6.2 Lake Tanganyika claroteines

The use of the LT claroteines as a comparative system was built upon the elucidation of the evolutionary relationships within this radiation. The novel molecular phylogeny presented in Chapter 2 reveals a double colonisation, by the ancestor of *Chrysicthys brachynema* and by the common ancestor of the remaining LT claroteines, followed by the diversification of the latter into the range of endemic genera now observed. These genera appear to have been diversifying at similar times, although there is not enough temporal resolution in the analysis to confidently attribute specific external factors as drivers of this diversification.

The species richness of the LT claroteines is shown to be currently underestimated. A greater number of potential *Phyllonemus* species than currently recorded is suggested both by morphology (R. Bills, pers. comm.) and, as demonstrated in Chapter 2, by genetic data. The presence of microallopatry is suggested as a potential driver for this underestimation of diversity, both in *Phyllonemus* where the proposed new species appear to be geographically restricted, and in *Lophiobagrus cyclurus*, which in Chapter 3 is shown to demonstrate a strong signal of phylogeographic structure within the lake.

The molecular phylogeny also highlights current problems with the taxonomy of the LT claroteines. *Chrysichthys* is shown not to be a monophyletic group, with the LT *Chrysichthys* clade more closely related to *Clarotes* than the non-LT *Chrysichthys* as previously thought (Bailey and Stewart, 1984). In addition, the position of *Chrysichthys sianenna,* which does not resolve within *Chrysichthys* or the LT *Chrysichthys,* is unclear. The instability of this taxon in analyses using

different molecular markers and methods of tree reconstruction indicates that further work, for example the sequencing of additional genetic markers, is required to elucidate the origin of this species.

In addition to providing insight into the evolutionary relationships of the LT claroteines, the molecular phylogeny also provides a basis from which to investigate morphological diversification through this radiation for the first time, as well as a basis from which to compare to the molecular phylogenies developed for the *Synodontis* radiation, and other endemic radiations.

6.3 Is there a common pattern of siluriform diversification in Lake Tanganyika?

The sympatric claroteine and *Synodontis* evolutionary radiations provide a valuable comparative dimension to investigate generalities in siluriform diversification in LT. In Chapters 3 and 4 the patterns and drivers of diversification seen in these two groups were investigated, both through a detailed intraspecific study of diversification patterns using two example species, and at a broader level by placing both radiations in a single phylogenetic context.

In Chapter 2, evidence of geographic separation was identified in multiple rocky shore claroteine species, including multiple *Phyllonemus* species, with support in the phylogeny for a similar pattern in *Lophiobagrus*. In Chapter 3, RAD-seq data was used to provide the first genomic level datasets in non-cichlid taxa from the East African Rift lakes, for *Lophiobagrus cyclurus*, their sister species *L. aquilus* and a comparator species *Synodontis multipunctatus*. These datasets were used to investigate phylogeographic structure in LT for the first time, in two separate and contrasting catfish lineages. The results showed that phylogeographic structure is important in some catfish species as evidenced by the strong signal in *Lophiobagrus cyclurus* including at very small spatial scales (78km in Zambia). This suggests that the same microallopatric patterns of diversification that are seen in some cichlids (e.g., Rico and Turner, 2002; Sefc et al., 2007; Van Oppen et al., 1997) are important in other taxa. However, just as in LT cichlids there are different patterns in different groups, as shown by the much weaker structure seen in *S. multipunctatus,* which is found a greater depths, 100m in this study but deeper in previous studies (Coulter, 1991), and shows different breeding behaviours.

By placing both the claroteine and *Synodontis* radiations is a common phylogenetic context, in Chapter 4, the relative adaptive nature of the two radiations could be investigated and contrasted. In the claroteine radiation each genus is distinct morphologically, indeed it is this morphological distinctness that resulted in their classification as different genera. Morphological measurements and stable isotope data show that in the claroteine radiation there is no significant overlap in morphology and diet between genera and an initial drop in subclade disparity provides evidence that this could be an adaptive radiation, as disparity is partitioned between subclades rather than within them. Changes in body shape appear to be linked to diet, for example more fusiform species display the isotopic signatures of a more pelagic diet. In contrast for *Synodontis* the opposite is true, with some evidence of convergence between subclades. LT *Synodontis* species converge on similar morphologies and diets with a greater width of dietary niche than is seen in the claroteines. This pattern is suggestive of a selective pressure to maintain dietary niche (stabilising selection rather than a divergent selective pressure leading to specialisation), however the presence of a hind gut in some species may well be an adaptation to diet. Although the overall patterns observed between the two complete radiations are dissimilar, the patterns seen within the individual genera of the claroteine radiation provide a useful comparative system to the *Synodontis* radiation.

6.4 How does diversification in Lake Tanganyika relate to patterns of diversification at the continental scale?

Synodontis is a widespread genus in Africa, with a continental distribution. By incorporating the insight gained on the LT *Synodontis* from the preceding chapters, Chapter 5 investigates the relationships between these lake scale patterns and those seen at a continental level. In concordance with Chapter 4, *Synodontis* in LT exhibit relatively little morphological change, with some evidence of convergence, however when these lake taxa are placed in a continental context with their congeners, taking into account branch lengths, we see that in fact the morphological divergence in LT is relatively large, given the small time scale, compared to *Synodontis* in riverine settings. However, the morphologies that have been developed in this in lacustrine setting, are not distinct to this environment and are similar to those found in riverine taxa, but with little evidence of explicit

convergence. A spike of diversification as a result of a change in selective regime following the colonisation of a new environment may be a more general pattern in *Synodontis* as a similar pattern is also observed in their recent colonisation of rivers in Southern Africa.

6.5 Future directions

In order to understand the factors driving diversification in LT and their relative roles in each radiation it is necessary to sample a broad range of different sized radiations (and taxa that have not radiated) in order to identify generalities between them. While the cichlids are the best studied system in terms of our understanding of the processes leading to diversification, and this thesis provides new evidence for the drivers of diversification in LT catfish radiations, the picture is far from complete. Additional studies within the catfish radiations, for example into breeding behaviour, would be useful in assessing the extent of on-going diversification, while additional genetic data would be useful to resolve uncertainty in the evolutionary relationships, in particular the placements of *Bathybagrus tetranema* and *Chrysichthys sianenna.* However, in order to investigate generalities between radiations it is necessary to investigate the drivers of diversification in additional radiations and particularly in the poorly studied non-vertebrate radiations. Although multiple such radiations have been identified (e.g., atyid prawns, Fryer, 2006; ostracods, Schön and Martens, 2011; gastropods, West and Michel, 2000; Wilson et al., 2004) only in the Platyphelphusid crabs has there been any investigation into the factors contributing to diversification with some evidence of niche partitioning (Marijnissen et al., 2008).

Studies into the catfish radiations are hampered by problems with the primary taxonomy. Many type specimens are old and some are unavailable, for example the type specimen for *C. graueri* (Steindachner, 1911) was unavailable for either of the LT *Chrysichthys* keys and this, coupled with different interpretations of the text, led to contradicting features in the keys of Bailey and Stewart (1984) and Hardman (2008). There is also taxonomic uncertainty within the LT *Synodontis* specimens. Multiple clades are resolved in all phylogenies (e.g., Chapter 4; Day and Wilkinson, 2006; Koblmüller et al., 2006) but the features used in the current key (Wright and Page, 2006) are not sufficient to delineate them, and

include multiple characters that do not preserve well, for example colouration, suggesting that additional characters would be useful. Both of these radiations represent the opportunity for traditional taxonomy to be aided by molecular phylogenies in producing accurate species descriptions. A taxonomic revision would also allow the evaluation of osteological characters in modern taxa that would aid in the placements of fossil taxa as these are often assigned to family using characters not used in modern taxa.

In combination with the problems in identifying taxa it is also clear that LT is under-sampled for catfishes with additional species to be described. Increased sampling, particularly from the Democratic Republic of Congo is important in order to assess species richness and distributions within LT. Increased fine scale sampling would be needed to investigate intraspecific patterns at smaller spatial scales than Chapter 3, reconstruct demographic expansions and assess the role of lake level fluctuations in catfishes which have been found to influence diversification in LT cichlids (Koblmüller et al., 2011; Nevado et al., 2013). An investigation into intraspecific geographic patterns in additional taxa with similar ecologies would provide evidence as to the generalities of geographic patterns. Further genomic resources would also enable investigation of the genetic changes underpinning morphological evolution or behaviours giving a more complete understanding of the process of diversification.

It can be problematic to place multiple catfish radiations in the same phylogeny as there is a lack of support for the relationships between major clades (Sullivan et al., 2006) and few accurately assigned fossils (discussed in Chapter 4) with most modern families described using soft tissue characters. However, the large siluriform phylogenies sampled at the genus level (Lundberg et al, 2007; Sullivan et al., 2006) were both based on the same dataset consisting of two exons of RAG1 and one from RAG2, suggesting that an increased number of genetic markers from the species in the phylogeny may increase the resolution and give more support to evolutionary relationships. This would enable further more distantly related catfish radiations to be put in the same phylogenetic context. For example, the endemic LT catfish radiations potentially represent a good comparative system to the catfish radiation in Lake Malawi, the *Bathyclarias-Clarias* radiation, for which the evolutionary relationships have been investigated

using mitochondrial data (Agnèse and Teugels, 2001) but there has been no investigation into the processes facilitating their diversification.

At a continental scale the role of palaeohydrology in driving diversification patterns in freshwater fish remains to be generalised with only *Synodontis* investigated throughout Africa (Day et al., 2013), with future sampling sites to extend geographical coverage for this genus suggested in Chapter 5. The Claroteinae would provide an interesting comparative system with which to explore palaeohydrological connections, however in order to do so increased sampling of claroteine species would be required, especially in the most speciose genus *Chrysichthys*. The Claroteinae are most diverse in the Congo basin, the majority of which falls in the Democratic Republic of Congo where the on-going political unrest hampers sampling effort. In addition to investigating the role of palaeohydrology, such sampling would also allow the identification of potential emigration of LT taxa into river systems and would include the riverine species *Chrysichthys dendrophorus,* which has no post cleithral process as seen in the LT taxa. Lakes have been suggested to be a source of riverine diversity in cichlids (Joyce et al., 2005), and *Synodontis victoriae* has been proposed to have emigrated from LT to riverine environments (Day and Wilkinson, 2006; Day et al., 2009; Day et al., 2013) but this has not yet been investigated in the Claroteinae.

Furthermore, full taxon sampling of the claroteine at continental scale would potentially allow ancestral range reconstruction as has been elucidated in *Synodontis* (Day et al., 2013) and mapping of museum collections may aid in understanding their present day distributions. A species level claroteine phylogeny coupled with morphological data would also provide an additional freshwater system with which to investigate if the same patterns between lineage accumulation and species diversification occur in the claroteine as those seen in *Synodontis*. The LT claroteine radiation occupies a greater area of morphospace than the LT *Synodontis* radiation suggesting they may have the potential to evolve greater morphological diversification.

6.6 References

Agnèse, J. F., & Teugels, G. G. (2001) The Bathyclarias-Clarias species flock: a new model to understand rapid speciation in African Great Lakes. *Comptes Rendus de l'Académie des Sciences - Series III - Sciences de la Vie,* **324**(8), 683–688

Bailey, R. M., & Stewart, D. J. (1984) Bagrid Catfishes from Lake Tanganyika, with a Key and Descriptions of New Taxa. *Miscellaneous Publications Museum of Zoology, University of Michigan*, **168**

Coulter, G. W. (1991) The benthic fish communities In G. W. Coulter (Ed.), *Lake Tanganyika and its Life.* Oxford: Oxford University Press. pp. 151–199

Day, J. J., Bills, R., & Friel, J. P. (2009) Lacustrine radiations in African Synodontis catfish. *Journal of Evolutionary Biology*, **22**(4), 805–817

Day, J. J., Peart, C. R., Brown, K. J., Friel, J. P., Bills, R., & Moritz, T. (2013) Continental diversification of an African catfish radiation (Mochokidae: Synodontis) *Systematic Biology*, **62**(3), 351-365

Day, J. J., & Wilkinson, M. (2006) On the origin of the Synodontis catfish species flock from Lake Tanganyika. *Biology Letters*, **2**(4), 548–552

Fryer, G. (2006) Evolution in ancient lakes: radiation of Tanganyikan atyid prawns and speciation of pelagic cichlid fishes in Lake Malawi. *Hydrobiologia*, **568**(S1), 131–142

Hardman, M. (2008) A new species of catfish genus Chrysichthys from Lake Tanganyika (Siluriformes: Claroteidae) *Copeia*, **2008**(1), 43

Joyce, D., Lunt, D., Bills, R., Turner, G., Katongo, C., Duftner, N., & Seehausen, O. (2005) An extant cichlid fish radiation emerged in an extinct Pleistocene lake. *Nature*, **435**, 90-95

Koblmüller, S., Salzburger, W., Obermüller, B., Eigner, E., Sturmbauer, C., & Sefc, K. M. (2011) Separated by sand, fused by dropping water: habitat barriers and fluctuating water levels steer the evolution of rock-dwelling cichlid populations in Lake Tanganyika. *Molecular Ecology*, **20**(11), 2272–2290

Koblmüller, S., Sturmbauer, C., Verheyen, E., Meyer, A., & Salzburger, W. (2006) Mitochondrial phylogeny and phylogeography of East African squeaker catfishes (Siluriformes: Synodontis) *BMC Evolutionary Biology*, **6**, 49

Lundberg, J. G., Sullivan, J. P., Rodiles-Hernández, R., & Hendrickson, D. A. (2007) Discovery of African roots for the Mesoamerican Chiapas catfish, Lacantunia enigmatica, requires an ancient intercontinental passage. *Proceedings of the Academy of Natural Sciences of Philadelphia,* **156**, 39–53

Marijnissen, S. A. E., Michel, E., Kamermans, M., Olaya-Bosch, K., Kars, M., Cleary, D. F. R., & Menken, S. B. J. (2008) Ecological correlates of species differences in the Lake Tanganyika crab radiation. *Hydrobiologia*, **615**(1), 81–94

Nevado, B., Mautner, S., Sturmbauer, C., & Verheyen, E. (2013) Water-level fluctuations and metapopulation dynamics as drivers of genetic diversity in populations of three Tanganyikan cichlid fish species. *Molecular Ecology*, **22**(15), 3933–3948

Rico, C., & Turner, G. F. (2002) Extreme microallopatric divergence in a cichlid species from Lake Malawi. *Molecular Ecology*, **11**(8), 1585–1590

Schön, I., & Martens, K. (2011) Molecular analyses of ostracod flocks from Lake Baikal and Lake Tanganyika. *Hydrobiologia*, **682**(1), 91–110

Sefc, K. M., Baric, S., Salzburger, W., & Sturmbauer, C. (2007) Species-specific population structure in rock-specialized sympatric cichlid species in Lake Tanganyika, East Africa. *Journal of Molecular Evolution*, **64**(1), 33–49

Steindachner, F. (1911) Beiträge zur Kenntniss der Fischfauna des Tanganyikasees und des Kongogebietes. *Anzeiger Der Akademie Der Wissenschaften in Wien*, **48**(26), 528–530

Sullivan, J. P., Lundberg, J. G., & Hardman, M. (2006) A phylogenetic analysis of the major groups of catfishes (Teleostei: Siluriformes) using rag1 and rag2 nuclear gene sequences. *Molecular Phylogenetics and Evolution*, **41**(3), 636–662

Van Oppen, M. J. H., Turner, G. F., Rico, C., Deutsch, J. C., Ibrahim, K. M., Robinson, R. L., & Hewitt, G. M. (1997) Unusually fine-scale genetic structuring found in rapidly speciating Malawi cichlid fishes. *Proceedings of the Royal Society B: Biological Sciences*, **264**(1389), 1803–1812

West, K., & Michel, E. (2000) The dynamics of endemic diversification: Molecular phylogeny suggests an explosive origin of the thiarid gastropods of Lake Tanganyika. In Rossiter, A., & Kawanabe, H. (Eds.) *Advances in Ecological Research* **31**. pp. 331–354

Wilson, A. B., Glaubrecht, M., & Meyer, A. (2004) Ancient lakes as evolutionary reservoirs: evidence from the thalassoid gastropods of Lake Tanganyika. *Proceedings of the Royal Society B: Biological Sciences*, **271**(1538), 529–536

Wright, J. J., & Page, L. M. (2006) Taxonomic revision of Lake Tanganyikan Synodontis (Siluriformes: Mochokidae) *Bulletin of the Florida Museum of Natural History*, **46**(4), 99–154

Appendix 1

Table 1 Table of specimens used in this study including sequence accession numbers. Institutional abbreviations are South African Institute of Aquatic Biodiversity (SAIAB) and Cornell University Museum of Vertebrates (CU).

Taxonomic level	Taxon	First described	Country	Locality	Water body	Field number	Vouch-ers	CO1	*Cytb*	RAG2	Plagl2	S7
Family	**Claroteidae**											
Subfamily	**Auchenoglanidinae**											
Genus	*Auchenoglanis*	Günther 1865										
Species	*Auchenoglanis occidentalis*	Valenciennes 1840	Zambia	Sumbu	Lake Tanganyika	C82, RB11B-043		HG803476	HG803392	HG803243	HG803316	HG803536
Species	*Auchenoglanis occidentalis*	Valenciennes 1840	Sudan	Kosti	White Nile	T347		HG803487	HG803403	HG803251		HG803542
Genus	*Parauchenoglanis*	Boulenger 1911										
Species	*Parauchenoglanis monkei*	Keilhack 1910	Benin	Igolo	Iguidi	T391		HG803488	HG803404	HG803252		
Species	*Parauchenoglanis monkei*	Keilhack 1910	Benin	Lokoli Forest	Iguidi	T568		HG803492	HG803408	HG803256	HG803325	HG803544
Species	*Parauchenoglanis ngamensis*	Boulenger 1911	Zambia		Sefula River		SAIAB 71869		HG803330	HG803190	HG803262	
Subfamily	**Claroteinae**	**Bleeker 1862**										
Genus	*Bathybagrus*	Bailey and Stewart 1984										
Species	*Bathybagrus tetranema*	Bailey and Stewart 1984	Zambia	Mpulungu	Lake Tanganyika	C206		HG803444	HG803360	HG803215	HG803287	HG803508
Species	*Bathybagrus tetranema*	Bailey and Stewart 1984	Zambia	Mpulungu	Lake Tanganyika	C287, RB11B-204		HG803454	HG803370		HG803297	
Species	*Bathybagrus tetranema*	Bailey and Stewart 1984	Zambia	Mpulungu	Lake Tanganyika	C364		HG803463	HG803379	HG803233	HG803306	HG803525

Taxonomic level	Taxon	First described	Country	Locality	Water body	Field number	Vouch-ers	CO1	*Cytb*	RAG2	Plagl2	S7
Genus	*Chrysichthys*	Bleeker 1858										
Species	*Chrysichthys ansorgii*	Boulenger 1910	Angola	Bengo	Kwanza River		SAIAB 84644	HG803428	HG803344	HG803201	HG803275	
Species	*Chrysichthys ansorgii*	Boulenger 1910	Angola	Bengo	Kawa River		SAIAB 84726	HG803430	HG803346			
Species	*Chrysichthys auratus*	Geoffroy Saint-Hilaire 1809	Sudan	Khartoum	White Nile	T270		HG803485	HG803401			
Species	*Chrysichthys auratus*	Geoffroy Saint-Hilaire 1809	Sudan	Kosti	White Nile	T337		HG803486	HG803402	HG803250	HG803321	
Species	*Chrysichthys* cf. *auratus*	Geoffroy Saint-Hilaire 1809	Benin		Pendjari National Park	T111		HG803482	HG803398			
Species	*Chrysichthys bocagii*	Boulenger 1910	Angola	Terra Nova village	Kwanza River		SAIAB 84622	HG803426	HG803342	HG803199	HG803273	
Species	*Chrysichthys bocagii*	Boulenger 1910	Angola	Cuanza Norte	Lucalla River		SAIAB 84685	HG803429	HG803345			
Species	*Chrysichthys brachynema*	Boulenger 1900	Tanzania	Kigoma	Lake Tanganyika	C3		HG803467	HG803383	HG803235	HG803308	HG803527
Species	*Chrysichthys brachynema*	Boulenger 1900	Zambia	Mpulungu	Lake Tanganyika		SAIAB 80287	HG803423	HG803339		HG803271	
Species	*Chrysichthys grandis*	Boulenger 1917	Tanzania	Kigoma	Lake Tanganyika		CU90324	HG803479	HG803395	HG803246		HG803539
Species	*Chrysichthys* cf. *grandis*	Boulenger 1917	Zambia	Mpulungu	Lake Tanganyika	C291		HG803455	HG803371	HG803225	HG803298	HG803518
Species	*Chrysichthys* cf. *grandis*		Zambia	Mpulungu	Lake Tanganyika	C207		HG803445	HG803361	HG803216	HG803288	HG803509
Species	*Chrysichthys mabusi*	Boulenger 1905	Zambia		Chambeshi River		SAIAB 118961	HG803412	HG803328	HG803188	HG803260	

Taxonomic level	Taxon	First described	Country	Locality	Water body	Field number	Vouch-ers	CO1	*Cytb*	RAG2	Plagl2	S7
Species	*Chrysichthys nigrodigitatus*	Lacepède 1803	Benin	Guezin	Lake Aheme	B88	SAIAB 75237	HG803417	HG803334	HG803193	HG803266	
Species	*Chrysichthys nigrodigitatus*	Lacepède 1803	Benin	Guezin	Lake Aheme	B89	SAIAB 75237	HG803416	HG803333	HG803192	HG803265	
Species	*Chrysichthys ornatus*	Boulenger 1902	Central African Republic	Dzanga-Sangha Forest	Sangha River		SAIAB 65201	HG803413	HG803329	HG803189	HG803261	
Species	*Chrysichthys platycephalus*	Worthington and Ricardo 1937	Tanzania	Kigoma	Lake Tanganyika	C1		HG803442	HG803358	HG803213	HG803285	HG803506
Species	*Chrysichthys platycephalus*	Worthington and Ricardo 1937	Zambia	Sumbu	Lake Tanganyika	C187, RB11B-070		HG803439	HG803355	HG803210	HG803282	HG803503
Species	*Chrysichthys platycephalus*	Worthington and Ricardo 1937	Tanzania	Kigoma	Lake Tanganyika	C5		HG803473	HG803389	HG803240	HG803313	HG803533
Species	*Chrysichthys platycephalus*	Worthington and Ricardo 1937	Zambia	Mpulungu	Lake Tanganyika	C335		HG803460	HG803376	HG803230	HG803303	HG803522
Species	*Chrysichthys sianenna*	Boulenger 1906	Zambia	Sumbu	Lake Tanganyika	C201		HG803443	HG803359	HG803214	HG803286	HG803507
Species	*Chrysichthys sianenna*	Boulenger 1906	Zambia	Mpulungu	Lake Tanganyika	C242		HG803451	HG803367	HG803222	HG803294	HG803515
Species	*Chrysichthys sianenna*	Boulenger 1906	Zambia	Mpulungu	Lake Tanganyika	C354		HG803462	HG803378	HG803232	HG803305	HG803524
Species	*Chrysichthys* sp.		Angola	Terra Nova village	Kwanza River		SAIAB 84627	HG803427	HG803343	HG803200	HG803274	
Species	*Chrysichthys* sp.		Angola		Kwanza River		SAIAB 84909	HG803431	HG803347	HG803202	HG803276	HG803496
Species	*Chrysichthys* sp.		Benin	Ouémé		T180		HG803483	HG803399	HG803249	HG803320	

Taxonomic level	Taxon	First described	Country	Locality	Water body	Field number	Vouch-ers	CO1	*Cytb*	RAG2	Plagl2	S7
Species	*Chrysichthys* sp.		Benin	Malanville		T244		HG803484	HG803400			
Species	*Chrysichthys* sp.		Benin	Tohonou	Lac Tohou	T47		HG803490	HG803406	HG803254	HG803323	
Species	*Chrysichthys* sp.		Benin	Ganvie	Cotonou Lagoon		SAIAB 75247	HG803418	HG803335	HG803194	HG803267	
Species	*Chrysichthys* sp.		Burkina Faso	Diebougou	Bougouriba River	T439		HG803489	HG803405	HG803253	HG803322	
Species	*Chrysichthys* sp.		Burkina Faso	Diebougou	Bougouriba River	T578		HG803494	HG803410	HG803258		
Species	*Chrysichthys* sp.		Tanzania	Kigoma	Lake Tanganyika		CU95204	HG803481	HG803397	HG803248		HG803541
Species	*Chrysichthys* sp.		Tanzania	Kigoma	Lake Tanganyika		CU95203	HG803480	HG803396	HG803247	HG803319	HG803540
Species	*Chrysichthys* sp.		Zambia	Mpulungu	Lake Tanganyika	C341		HG803461	HG803377	HG803231	HG803304	HG803523
Species	*Chrysichthys* sp.		Zambia	Mpulungu	Lake Tanganyika		SAIAB 76284	HG803420	HG803337	HG803195	HG803268	HG803495
Species	*Chrysichthys* sp.		Zambia	Mpulungu	Lake Tanganyika	C292		HG803456	HG803372	HG803226	HG803299	HG803519
Genus	*Clarotes*	Kner 1855										
Species	*Clarotes laticeps*	Rüppell 1829	Benin	Malanville		T539		HG803491	HG803407	HG803255	HG803324	HG803543
Species	*Clarotes laticeps*	Rüppell 1829	Benin		Pendjari River	T576		HG803493	HG803409	HG803257	HG803326	HG803545
Genus	*Lophiobagrus*	Poll 1942										
Species	*Lophiobagrus aquilus*	Bailey and Stewart 1984	Zambia	Kombe	Lake Tanganyika		SAIAB 80316	HG803425	HG803341	HG803198	HG803272	
Species	*Lophiobagrus aquilus*	Bailey and Stewart 1984	Zambia	Sumbu	Lake Tanganyika	C112		HG803435	HG803351	HG803206		HG803499
Species	*Lophiobagrus aquilus*	Bailey and Stewart 1984	Zambia	Sumbu	Lake Tanganyika	C119		HG803437	HG803353	HG803208	HG803280	HG803501

Taxonomic level	Taxon	First described	Country	Locality	Water body	Field number	Vouch-ers	CO1	*Cytb*	RAG2	Plagl2	S7
Species	*Lophiobagrus aquilus*	Bailey and Stewart 1984	Zambia	Mpulungu	Lake Tanganyika	C213		HG803446	HG803362	HG803217	HG803289	HG803510
Species	*Lophiobagrus aquilus*	Bailey and Stewart 1984	Zambia	Mpulungu	Lake Tanganyika	C238		HG803449	HG803365	HG803220	HG803292	HG803513
Species	*Lophiobagrus brevispinis*	Bailey and Stewart 1984	Zambia	Mpulungu	Lake Tanganyika	C222		HG803448	HG803364	HG803219	HG803291	HG803512
Species	*Lophiobagrus brevispinis*	Bailey and Stewart 1984	Zambia	Mpulungu	Lake Tanganyika	C319		HG803458	HG803374	HG803228	HG803301	
Species	*Lophiobagrus brevispinis*	Bailey and Stewart 1984	Zambia	Sumbu	Lake Tanganyika	C85		HG803477	HG803393	HG803244	HG803317	HG803537
Species	*Lophiobagrus cyclurus*	Worthington and Ricardo 1937	Burundi		Lake Tanganyika	C379		HG803464	HG803380	HG803234	HG803307	
Species	*Lophiobagrus cyclurus*	Worthington and Ricardo 1937	Burundi		Lake Tanganyika	C398		HG803466	HG803382			
Species	*Lophiobagrus cyclurus*	Worthington and Ricardo 1937	Tanzania	Kigoma	Lake Tanganyika	C4		HG803471	HG803387	HG803239	HG803312	HG803531
Species	*Lophiobagrus cyclurus*	Worthington and Ricardo 1937	Tanzania	Kigoma	Lake Tanganyika	C53		HG803472	HG803388			HG803532
Species	*Lophiobagrus cyclurus*	Worthington and Ricardo 1937	Tanzania	Kigoma	Lake Tanganyika	C10		HG803433	HG803349	HG803204		HG803497
Species	*Lophiobagrus cyclurus*	Worthington and Ricardo 1937	Tanzania	Kigoma	Lake Tanganyika	C37		HG803465	HG803381			HG803526
Species	*Lophiobagrus cyclurus*	Worthington and Ricardo 1937	Zambia	Mpulungu	Lake Tanganyika	C244		HG803452	HG803368	HG803223	HG803295	HG803516

Taxonomic level	Taxon	First described	Country	Locality	Water body	Field number	Vouch-ers	CO1	*Cytb*	RAG2	Plagl2	S7
Species	*Lophiobagrus cyclurus*	Worthington and Ricardo 1937	Zambia	Mpulungu	Lake Tanganyika		SAIAB 76161	HG803419	HG803336			
Genus	*Phyllonemus*	Boulenger 1906										
Species	*Phyllonemus* aff. *brichardi*		Zambia	Sumbu	Lake Tanganyika	C111		HG803434	HG803350	HG803205	HG803278	HG803498
Species	*Phyllonemus* aff. *brichardi*		Zambia	Sumbu	Lake Tanganyika	C71		HG803474	HG803390	HG803241	HG803314	HG803534
Species	*Phyllonemus* aff. *brichardi*		Zambia	Sumbu	Lake Tanganyika	C76		HG803475	HG803391	HG803242	HG803315	HG803535
Species	*Phyllonemus filinemus*	Worthington and Ricardo 1937	Tanzania	Kigoma	Lake Tanganyika	C2		HG803457	HG803373	HG803227	HG803300	HG803520
Species	*Phyllonemus filinemus*	Worthington and Ricardo 1937	Tanzania	Kigoma	Lake Tanganyika	C21		HG803447	HG803363	HG803218	HG803290	HG803511
Species	*Phyllonemus filinemus*	Worthington and Ricardo 1937	Tanzania	Kigoma	Lake Tanganyika	C42		HG803469	HG803385	HG803237	HG803310	HG803529
Species	*Phyllonemus typus*	Boulenger 1906	Zambia	Sumbu	Lake Tanganyika	C136		HG803438	HG803354	HG803209	HG803281	HG803502
Species	*Phyllonemus typus*	Boulenger 1906	Zambia	Sumbu	Lake Tanganyika	C193, RB11B-069		HG803441	HG803357	HG803212	HG803284	HG803505
Species	*Phyllonemus typus*	Boulenger 1906	Zambia	Mpulungu	Lake Tanganyika	C324		HG803459	HG803375	HG803229	HG803302	HG803521
Species	*Phyllonemus* sp. A		Burundi	Gitaza	Lake Tanganyika		SAIAB 75112	HG803414	HG803331	HG803191	HG803263	
Species	*Phyllonemus* sp. A		Burundi	Kagongo	Lake Tanganyika		SAIAB 75149	HG803415	HG803332		HG803264	

Taxonomic level	Taxon	First described	Country	Locality	Water body	Field number	Vouch-ers	CO1	*Cytb*	RAG2	Plagl2	S7
Species	*Phyllonemus* sp. A		Burundi		Lake Tanganyika	C415		HG803468	HG803384	HG803236	HG803309	HG803528
Species	*Phyllonemus* sp. B		Tanzania	Kigoma	Lake Tanganyika	C23		HG803450	HG803366	HG803221	HG803293	HG803514
Species	*Phyllonemus* sp. B		Tanzania	Kigoma	Lake Tanganyika	C24		HG803453	HG803369	HG803224	HG803296	HG803517
Species	*Phyllonemus* sp. B		Tanzania	Kigoma	Lake Tanganyika	C44		HG803470	HG803386	HG803238	HG803311	HG803530
Species	*Phyllonemus* sp. C		Zambia	Sumbu	Lake Tanganyika	C115		HG803436	HG803352	HG803207	HG803279	HG803500
Species	*Phyllonemus* sp. C		Zambia	Sumbu	Lake Tanganyika	C190		HG803440	HG803356	HG803211	HG803283	HG803504
Species	*Phyllonemus* sp. C		Zambia	Sumbu	Lake Tanganyika	C92		HG803478	HG803394	HG803245	HG803318	HG803538
Species	*Phyllonemus* sp. D		Zambia	Mpulungu	Lake Tanganyika		SAIAB 80289	HG803424	HG803340			
Species	*Phyllonemus* sp. D		Zambia	Mpulungu	Lake Tanganyika		SAIAB 118960	HG803411	HG803327	HG803187	HG803259	
Family	**Schilbeidae**											
Genus	Parailia	Boulenger 1899										
Species	*Parailia congica*	Boulenger 1899	Central African Republic		Baidou		SAIAB 77672	HG803421		HG803196	HG803269	
Genus	*Pareutropius*	Regan 1920										
Species	*Pareutropius debauwi*	Boulenger 1900	Central African Republic		Baidou		SAIAB 77676	HG803422	HG803338	HG803197	HG803270	
Genus	*Schilbe*	Oken 1817										

Taxonomic level	Taxon	First described	Country	Locality	Water body	Field number	Vouch-ers	CO1	*Cytb*	RAG2	Plagl2	S7
Species	*Schilbe intermedius*	Rüppell 1832	Mozambique	Tete			SAIAB 97045	HG803432	HG803348	HG803203	HG803277	

Appendix 2

Table 1 Sample number, species, location and GPS coordinates for all specimens used in this study

Sample No.	Species Name	Location	Latitude	Longitude	Barcode	Library
C9	*Lophiobagrus cyclurus*	Kigoma	-4.880194	29.620944	AAAAA	2
C12	*Lophiobagrus cyclurus*	Kigoma	-4.880194	29.620944	TTAAT	4
C15	*Lophiobagrus cyclurus*	Kigoma	-4.886472	29.612861	TGGTT	4
C34	*Lophiobagrus cyclurus*	Kigoma	-4.880194	29.620944	CGATA	3
C36	*Lophiobagrus cyclurus*	Kigoma	-4.880194	29.620944	AAGGG	2
C38	*Lophiobagrus cyclurus*	Kigoma	-4.880194	29.620944	CTAGG	3
C40	*Lophiobagrus cyclurus*	Kigoma	-4.880194	29.620944	TGACC	4
C41	*Lophiobagrus cyclurus*	Kigoma	-4.880194	29.620944	CGGCG	3
C113	*Lophiobagrus aquilus*	Sumbu	-8.483306	30.467333	TAATG	3
C119	*Lophiobagrus aquilus*	Sumbu	-8.483306	30.467333	TAGCA	3
C134	*Lophiobagrus cyclurus*	Sumbu	-8.476056	30.449444	ATATC	4
C141	*Lophiobagrus cyclurus*	Sumbu	-8.476056	30.449444	GGGGA	3
C161	*Lophiobagrus cyclurus*	Sumbu	-8.421222	30.457778	GTACA	3
C166	*Lophiobagrus cyclurus*	Sumbu	-8.421222	30.457778	GTGTG	3
C171	*Lophiobagrus cyclurus*	Sumbu	-8.421222	30.457778	CAGTC	2
C172	*Lophiobagrus cyclurus*	Sumbu	-8.418417	30.461417	ACGTA	4
C174	*Lophiobagrus cyclurus*	Sumbu	-8.418417	30.461417	AAGGG	4
C175	*Lophiobagrus cyclurus*	Sumbu	-8.418417	30.461417	ACACG	4
C227	*Lophiobagrus cyclurus*	Mpulungu	-8.752972	31.084667	GCGCC	3
C228	*Lophiobagrus aquilus*	Mpulungu	-8.752972	31.084667	AGGAC	4
C229	*Lophiobagrus cyclurus*	Mpulungu	-8.752972	31.084667	ACCAT	4
C230	*Lophiobagrus cyclurus*	Mpulungu	-8.752972	31.084667	GCATT	3
C236	*Lophiobagrus aquilus*	Mpulungu	-8.752972	31.084667	CCAAC	2
C237	*Lophiobagrus cyclurus*	Mpulungu	-8.752972	31.084667	ACTGC	4
C238	*Lophiobagrus aquilus*	Mpulungu	-8.752972	31.084667	TCAGA	3
C243	*Lophiobagrus cyclurus*	Mpulungu	-8.753278	31.098222	AAAAA	4
C245	*Lophiobagrus cyclurus*	Mpulungu	-8.753278	31.098222	ATGCT	4
C246	*Lophiobagrus cyclurus*	Mpulungu	-8.753278	31.098222	GGAAG	3
C261	*Lophiobagrus cyclurus*	Mpulungu	-8.753278	31.098222	AGAGT	2
C309	*Lophiobagrus aquilus*	Mpulungu	-8.797472	31.019667	TCGAG	3
C311	*Lophiobagrus aquilus*	Mpulungu	-8.797472	31.019667	AGAGT	4
C365	*Lophiobagrus cyclurus*	Burundi	-3.675278	29.334167	GAAGC	3
C373	*Lophiobagrus cyclurus*	Burundi	-3.675278	29.334167	TTGGC	4
C378	*Lophiobagrus cyclurus*	Burundi	-3.675278	29.334167	ACACG	2
C379	*Lophiobagrus cyclurus*	Burundi	-3.675278	29.334167	CTGAA	3
C383	*Lophiobagrus cyclurus*	Burundi	-3.675278	29.334167	GAGAT	3
C389	*Lophiobagrus cyclurus*	Burundi	-3.675278	29.334167	AACCC	4
C398	*Lophiobagrus cyclurus*	Burundi	-3.675278	29.334167	ACGTA	2
C406	*Lophiobagrus cyclurus*	Burundi	-3.675278	29.334167	AATTT	4
S2	*Synodontis multipunctatus*	Kigoma	-4.886472	29.612861	CGATA	1
S3	*Synodontis multipunctatus*	Kigoma	-4.886472	29.612861	CGGCG	1
S28	*Synodontis multipunctatus*	Kigoma	-4.880194	29.620944	GAGAT	1
S29	*Synodontis multipunctatus*	Kigoma	-4.880194	29.620944	GAAGC	1

Sample No.	Species Name	Location	Latitude	Longitude	Barcode	Library
S35	*Synodontis multipunctatus*	Kigoma	-4.880194	29.620944	CTGAA	1
S38	*Synodontis multipunctatus*	Kigoma	-4.886472	29.612861	GCGCC	1
S40	*Synodontis multipunctatus*	Kigoma	-4.886472	29.612861	GCATT	1
S41	*Synodontis multipunctatus*	Kigoma	-4.886472	29.612861	CTAGG	1
S160	*Synodontis multipunctatus*	Mpulungu	-8.752972	31.084667	TGACC	2
S275	*Synodontis multipunctatus*	Mpulungu	-8.743667	31.059694	TGGTT	1
S176	*Synodontis multipunctatus*	Mpulungu	-8.743667	31.059694	AACCC	2
S180	*Synodontis multipunctatus*	Mpulungu	-8.743667	31.059694	AATTT	2
S187	*Synodontis multipunctatus*	Mpulungu	-8.753278	31.098222	ACCAT	2
S205	*Synodontis multipunctatus*	Mpulungu	-8.751528	31.033222	ACTGC	2
S221	*Synodontis multipunctatus*	Mpulungu	-8.751528	31.033222	TTAAT	2
S249	*Synodontis multipunctatus*	Mpulungu	-8.767139	31.099056	TTGGC	2
S254	*Synodontis multipunctatus*	Burundi	-3.616149	29.343914	TCAGA	1
S257	*Synodontis multipunctatus*	Burundi	-3.616149	29.343914	TCGAG	1
S270	*Synodontis multipunctatus*	Burundi	-3.616149	29.343914	TAGCA	1
S175	*Synodontis multipunctatus*	Burundi	-3.616149	29.343914	GGGGA	3
S282	*Synodontis multipunctatus*	Burundi	-3.616149	29.343914	GTACA	1
S285	*Synodontis multipunctatus*	Burundi	-3.616149	29.343914	GTGTG	1
S288	*Synodontis multipunctatus*	Burundi	-3.616149	29.343914	GGAAG	1
S290	*Synodontis multipunctatus*	Burundi	-3.616149	29.343914	TAATG	1

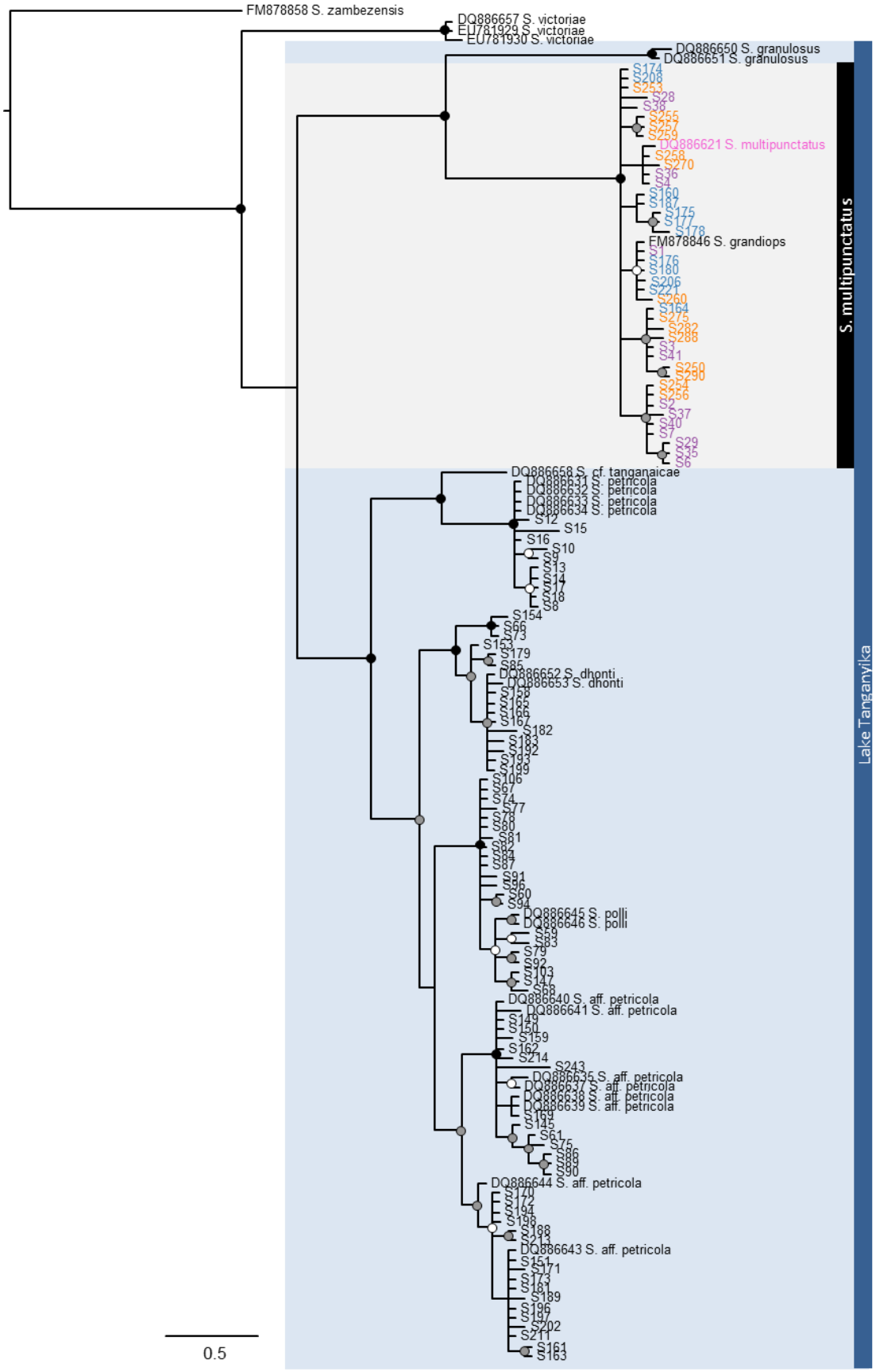


Figure 1 Bayesian tree constructed using MrBayes (Huelsenbeck and Ronquist, 2001) using *Cytb* data for Lake Tanganyika *Synodontis* using novel sequences (shown as S###) and previously published sequences from GenBank (shown with accession no. and species name). *Synodontis multipunctatus* are coloured by location (blue : Mpulungu, purple : Kigoma, orange: Burundi, pink : Democratic Republic of Congo). Circles on nodes represent posterior probability (black : 1, grey : ≥ 0.95, white : ≥ 0.9, no circle : <0.9).

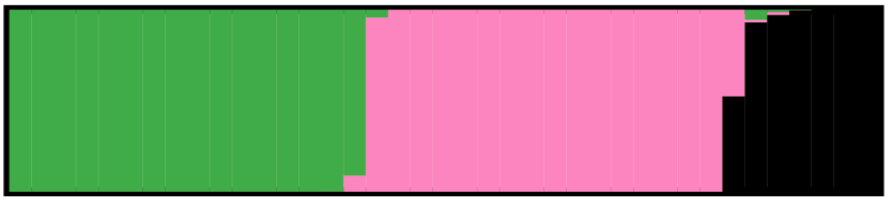


Figure 2 STRUCTURE Plot for *Lophiobagrus cyclurus* and *Lophiobagrus aquilus* (K=3). Green in the STRUCTURE plots represents *L. cyclurus* from both sites from the northern basin, pink, *L. cyclurus* from the southern basin, and black *L. aquilus*.

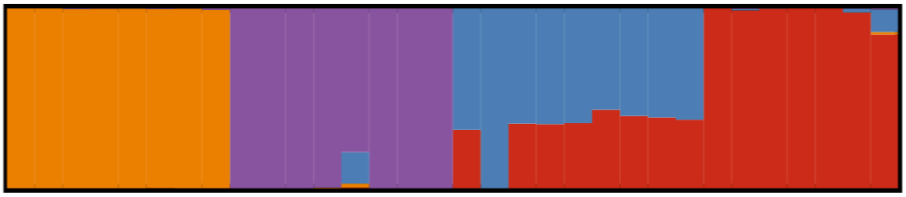


Figure 3 STRUCTURE Plot for *Lophiobagrus cyclurus* only (K=4). Orange in the STRUCTURE plots represents samples from Burundi, purple, samples from Kigoma, blue, samples from Mpulungu and red samples from Sumbu.

Figure 4 STRUCTURE Plot for *Lophiobagrus cyclurus* only (K=3). Orange in the STRUCTURE plots represents samples from Burundi, purple, samples from Kigoma, and yellow from the southern basin.

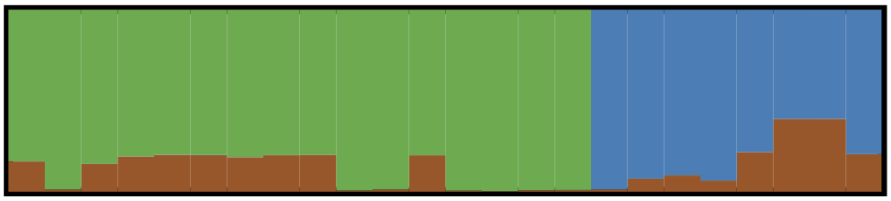


Figure 5 STRUCTURE Plot for *Synodontis multipunctatus* (K=3). Green represents the northern basin, blue, Mpulungu and brown a third cluster observed across samples from both locations.

Table 2 Fst values for *Lophiobagrus* as calculated in the python library egglib

	Burundi	**Kigoma**	**Sumbu**	**Mpulungu**
Kigoma	0.2020			
Sumbu	0.1594	0.2015		
Mpulungu	0.1664	0.2044	0.0330	
L. aquilus	0.4011	0.4917	0.3355	0.3782

Table 3 Fst values for *S. multipunctatus* as calculated in the python library egglib

	Burundi	**Kigoma**
Kigoma	-0.0195	
Mpulungu	-0.0008	0.0009

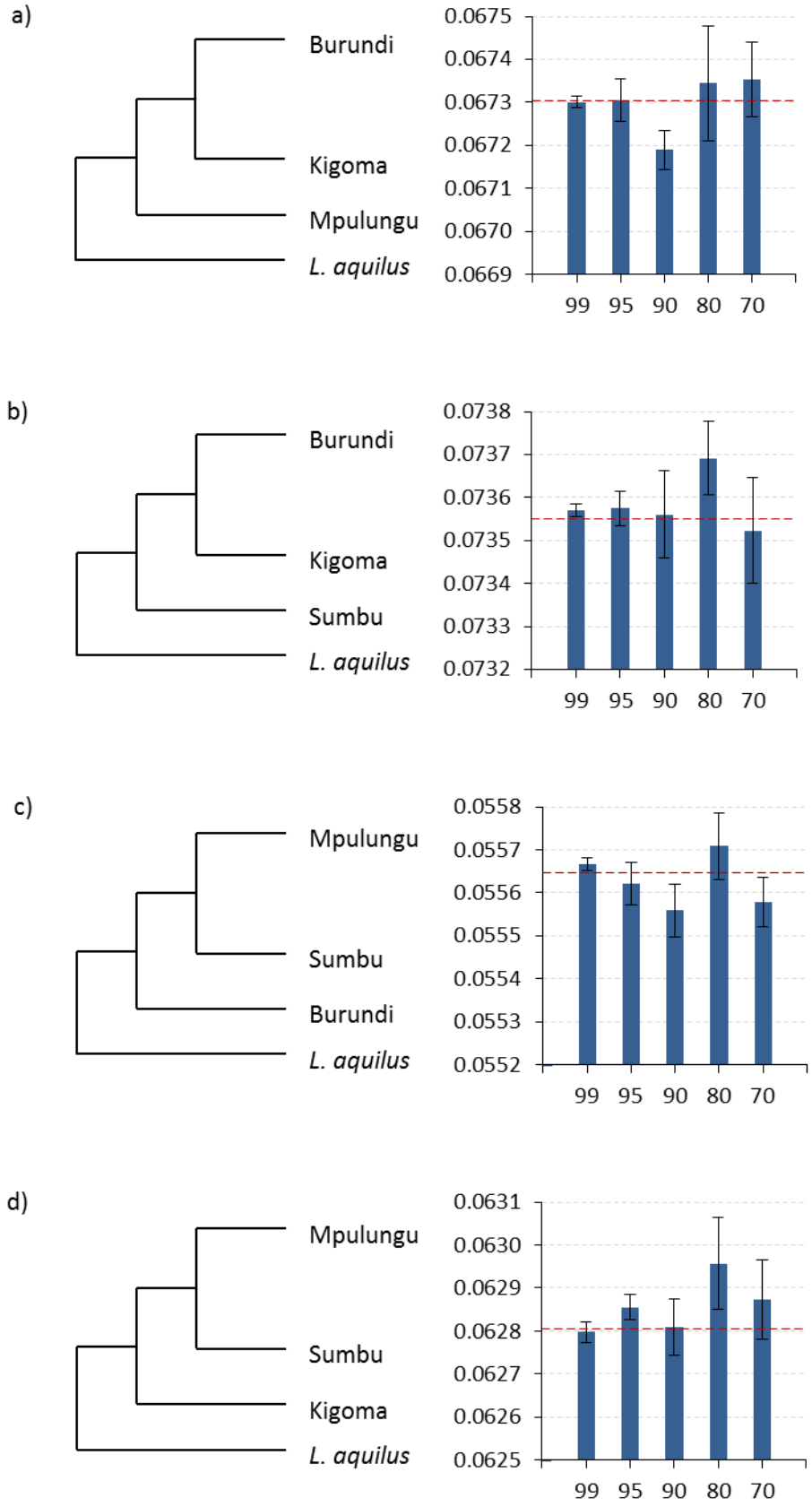


Figure 6 Nominal D Statistics for ABBA-BABA test for Lophiobagrus, without sample C228. Red line is the overall nominal D value for each topology. Blue bars represent the mean nominal D value from 1,000 random subsamples of the dataset at each percentage coverage. Error bars are the standard deviation (standard error of the mean is not visible at this scale).

Appendix 3

Table 1 List of specimens from Lake Tanganyika used in the morphological and stable isotope analyses in this study.

Species	Morphological Analysis	Stable Isotope Analysis
Bathybarus tetranema	RMCA83-04-P-1_2, RMCA83-04-P-1_2, RMCA94-031-P-0026, RMCA95-098-P-0044-0050, RMCA95-098-P-0044-0050, RMCA95-098-P-0044-0050, RMCA95-098-P-0044-0050, RMCA95-098-P-0041-0043, RMCA95-098-P-0041-0043, RMCA95-098-P-0041-0043	C287, C289, C364
Chrysichthys acsiorum	AMNH236052, AMNH217411, AMNH217411, AMNH217411	
Chrysichthys grandis	CU90324, RMCA14347, RMCA94-069-P-0216-0218	C291
Chrysichthys graueri	CU95203 (JPF 1627), CU95203 (JPF 1626), RMCA96-083-P-0685-0687, RMCA128678, RMCA95-098-P-0066	C292
Chrysichthys platycephalus	CU95204 (JPF 1624), CU95204 (JPF 1625), RMCA92-081-P-0167, RMCA92-081-P-0169, RMCA63791-63792, RMCA63791-63792, RMCA44994-44996, RMCA44994-44996, RMCA95-098-P-0067-0070, RMCA95-098-P-0067-0070, RMCA83-19-P-3, RMCA91-034-P-0620, C33, C14, C56, CU88726 (213), CU88726 (203)	C105, C107, C108, C110, C131, C156, C158, C187, C189, C240, C262, C290, C293, C294, C315, C335, C345, C362, C363, C65
Chrysichthys sianenna	RMCA92-081-P-0146, RMCA92-081-P-0104, RMCA92-081-P-1659, RMCA92-081-P-1785-1800, RMCA92-081-P-1660-1667, RMCA92-081-P-1660-1667, RMCA92-081-P-1660-1667, RMCA92-081-P-1660-1667, RMCA92-081-P-1660-1667, RMCA92-081-P-1660-1667, AMNH217384, AMNH97210	C199, C201, C241, C242, C288, C314, C354, C75
Chrysichthys stappersii	RMCA90189, RMCA14236	
Lophiobagrus aquilus	C228, C311, C238, C73, C236, RMCA94-031-P-0034, RMCA83-04-P-3-7, RMCA83-04-P-3-7, C112, C66	C228, C311, C238, C73, C236, C113, C119, C211, C213, C309, C322, C326
Lophiobagrus asperispinis	RMCA14359A, RMCA14359B, RMCA92-081-P-1677, BMNH 1920-5-25-75	
Lophiobagrus brevispinis	RMCA131093, RMCA81-16-P-1-13, RMCA81-16-P-1-13, RMCA81-16-P-1-13, RMCA81-16-P-1-13, RMCA81-16-P-1-13, RMCA131093, BMNH 1983-2-8:7-10, BMNH 1983-2-8:7-10, BMNH 1983-2-8:7-10	C114, C117, C123, C125, C127, C130, C192, C214, C219, C221, C222, C232, C249, C252, C253, C254, C296, C77, C83, C85, C94, C98
Lophiobagrus cyclurus (B)	C398, C379, C408, C365, C378, C384, C383, C375, C381	

Species	Morphological Analysis	Stable Isotope Analysis
Lophiobagrus cyclurus (K)	C15, C40, C38, C9, C41, C53, C36, C4, C38, C40, C389, C10	
Lophiobagrus cyclurus (Z)	C155, C139, C149, C243, C268, C133, C173, C264, C134, C267	C155, C139, C149, C243, C268, C137, C138, C146, C148, C150, C154, C159, C210, C233, C239, C244, C245, C260, C261, C265
Phyllonemus aff. Brichardi	C111	C111, C71, C76
Phyllonemus filinemus	RMCA92-081-P-0141, RMCA92-081-P-1678, C22, C21, C6, C19, C2, C42, C26, C49	
Phyllonemus sp. B	C23, C24, C44, C17, C18, C16, C20, C29, C25, C28, C45	
Phyllonemus sp. C	C61, C92, C190, C95, C115, C96	C61, C92, C190, C95, C115, C96, C118, C91, C97, C99
Phyllonemus typus	C144, C324, C145, C136, C188, C147, RMCA90250-90252, RMCA90249	C144, C324, C145, C136, C188, C147, C132, C140, C193, C194, C195, C323
Clade 1	CU88758	
Clade 2	BMNH 2006.3.6.16 (5208), BMNH 2006.3.6.15(5126), BMNH 2006.3.6.18 (5046), BMNH 2006.3.6.17 (5213), S9, S14, S17, S8, S15, S12, S10, S18, S13	
Clade 3	S183, S153, S164, S166, S154, S199, S165, S167, S158, S167, BMNH 2005-9-26-3 (05148), S179	S183, S153, S164, S166, S154, S199, S165, S167, S158, S167, S182, S192, S193, S66, S73, S85
Clade 4	S78, S106, S94, S59, S81, S80, S60, S83, S103, BMNH 2005-9-26-18 (5052), BMNH 2005-9-26-2 (5100), S150	S78, S106, S94, S59, S81, S80, S60, S83, S103, S147, S67, S68, S74, S77, S79, S82, S84, S87, S91, S92, S96
Clade 5	S214, S243, S162, S159, BMNH 2006-3-6-30 (5149), BMNH 2006-3-6-29 (5145), BMNH 2006-3-6-31 (5146), BMNH 2006-3-6-32 (5147), BMNH 2007-8-29-28-30(5152), S149, S150, S149	S214, S243, S162, S159, S145, S169, S61, S75, S86, S89, S90
Clade 6	S173, S198, S172, S213, S211, S163, S197, S196, S173, S170, S161, S171, BMNH 2005-9-26-1 (5124), BMNH 2007-8-29-28-30(5153)	S173, S198, S172, S213, S211, S163, S197, S196, S173, S170, S161, S171, S151, S181, S188, S189, S194, S202
Synodontis dhonti	14344	
Synodontis grandiops	BMNH 1982-4-13-4785, BMNH 1982-4-13-4784, BMNH 1982-4-13-4789-4791 (3), BMNH 1982-4-13-4789-4791 (4), BMNH 1982-4-13-4789-4791 (1), BMNH 1982-4-13-4786, BMNH 1955-12-20-1837, BMNH 1955-12-20-1833, BMNH 1982-4-13-4787-4788 (2), BMNH 1982-4-13-4787-4788 (1)	
Synodontis granulosus	82-12-P-13-16, 82-12-P-13-16, 82-12-P-13-16, 82-12-P-13-16, 94-069-P-0289, A1-094-P-0052, 100902, 14157, BMNH 1906-9-6-40, BMNH 1936-6-15-1199-1201	

Species	Morphological Analysis	Stable Isotope Analysis
Synodontis multipunctatus	BMNH 2005-9-26-19-23, BMNH 2005-9-26-19-23, S3, S35, S2, S288, S285, S257, S282, S270, S275, S254	S160, S174, S175, S176, S177, S178, S180, S187, S205, S206, S207, S208, S221, S249

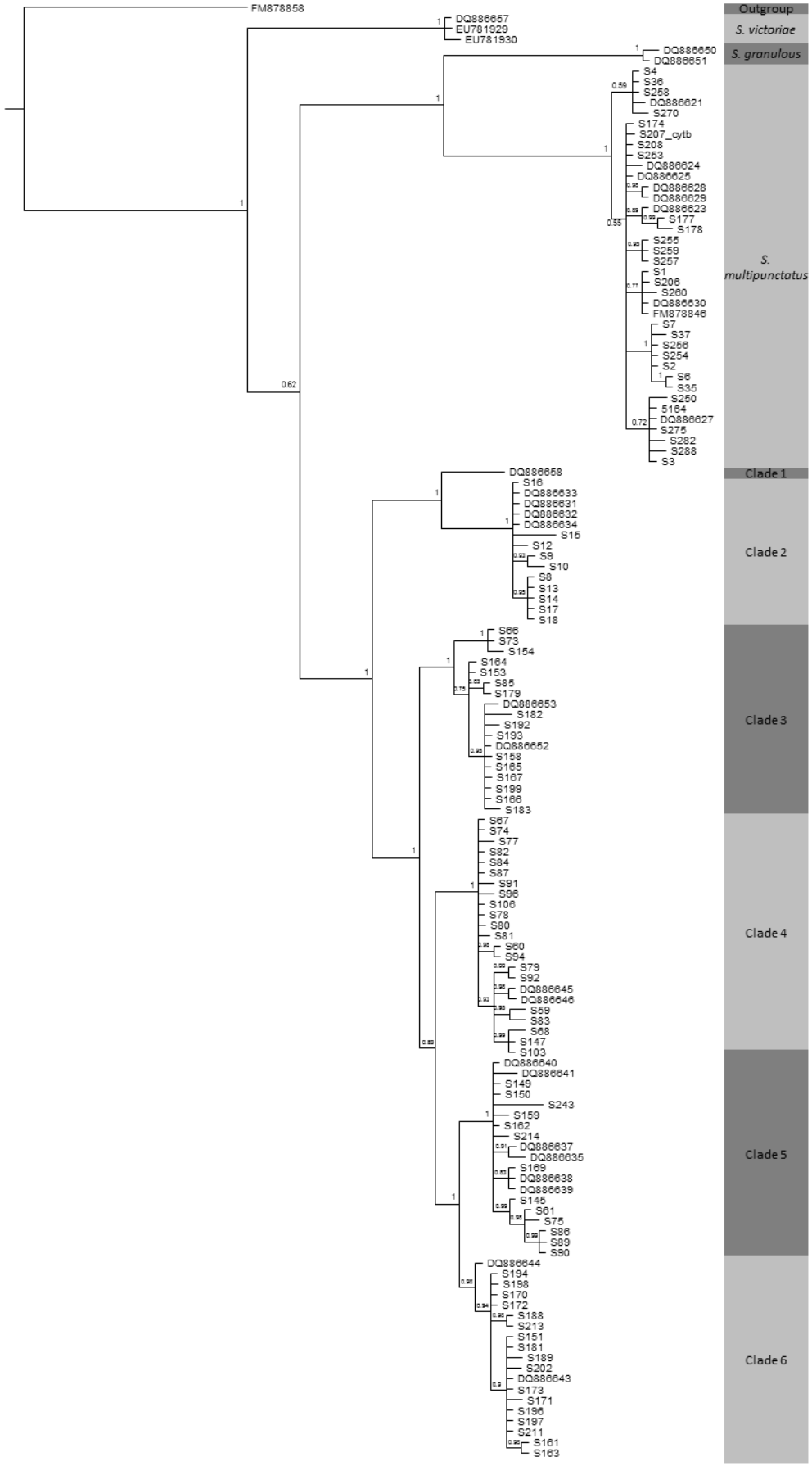


Figure 1 Bayesian *Cytb* tree showing the clade designations for *Syndodontis* samples used in this study.

Table 2 List of genetic samples used in the Ostariophysian and 'Big Africa' phylogenies. Novel sequences created for this study are marked as 'Novel', sequences generated in the laboratory of Dr. Thomas Near are marked 'Near', sequences present in GenBank are marked with their GenBank accession number.

Species	RAG1 Exon 3	ENC1	Plagl2	RAG2	CO1	*Cytb*
Acanthodoras cataphractus	DQ492466					
Acrochordonichthys rugosus	DQ492444					
Ageneiosus ucayalensis	DQ492463					
Ailia coila	DQ492452					
Akysis sp.	DQ492445					
Alosa pseudoharengus		Near	Near			
Amblyceps sp.	DQ492451					
Ameiurus natalis	Near	Near	Near			
Ameiurus nebulosus	DQ492510					
Amphilius cf. jacksonii	Near	Near	Near	DQ492378		
Amphilius uranoscopus	Novel	Novel	Novel	Novel		
Anaspidoglanis macrostoma	DQ492499			DQ492386		
Anduzedoras oxyrhynchus	Near	Near	Near			
Apteronotus albifrons	Near	Near	Near			
Arius felis	Near	Near	Near			
Astroblepus sp. 1	DQ492438					
Astroblepus sp. 2	DQ492439					
Astyanax mexicanus		Near	Near			
Atopochilus savorgnani	DQ492493			DQ492380		
Auchenoglanis occidentalis	Novel	Novel		HG803251	HG803487	HG803403
Bagarius yarrelli	DQ492446					
Bagre marinus	DQ492524					
Bagrichthys macropterus	Near	Near	Near			
Bagrus docmak	Near	Near	Near			
Bagrus ubangensis	Near	Near	Near			
Barbatula barbatula	Near		Near			
Batasio tigrinus	DQ492460					
Bathybagrus tetranema	DQ492502		HG803287	HG803215	HG803444	HG803360
Batrochoglanis raninus	DQ492473					
Belonoglanis sp.	Novel		Novel	Novel		
Belonoglanis tenuis	DQ492489			DQ492376		
Brachyplatystoma filamentosum	Near	Near	Near			

Species	RAG1 Exon 3	ENC1	Plagl2	RAG2	CO1	*Cytb*
Brycon pesu	Near	Near	Near			
Bullockia maldonadoi	DQ492434					
Callichthys callichthys	Near	Near	Near			
Carpiodes carpio	Near		Near			
Catostomus commersoni	Near	Near	Near			
Centromochlus heckelii	DQ492465					
Cephalocassis borneensis	DQ492525					
Cetopsis candiru	DQ492533					
Cetopsis coecutiens	Near		Near			
Chaca chaca	DQ492469					
Chaca sp.	DQ492470					
Chalceus macrolepidotus	Near	Near	Near			
Chanos chanos	Near	Near	Near			
Chiloglanis niloticus	Novel	Novel	Novel	HF565738	HF565846	HF565994
Chrysichthys auratus	Novel	Novel	HG803321	HG803250	HG803486	HG803402
Chrysichthys brachynema	Novel		HG803308	HG803235	HG803467	HG803383
Chrysichthys mabusi	Novel		HG803260	HG803188	HG803412	HG803328
Chrysichthys platycephalus	Novel	Novel	HG803282	HG803210	HG803439	HG803355
Chrysichthys sianenna	Novel	Novel	HG803286	HG803214	HG803443	HG803359
Chrysichthys sp.	Novel		HG803288	HG803216	HG803445	HG803361
Chrysichthys sp.	Novel		HG803304	HG803231	HG803461	HG803377
Citharinus congicus	Near		Near			
Clarias batrachus	DQ492521					
Clarias gabonensis	DQ492519					
Clarotes laticeps	Novel		HG803324	HG803255	HG803491	HG803407
Conorhynchos conirostris	DQ492477					
Corydoras aurofrenatus	Near	Near	Near			
Corydoras cf. trilineatus	DQ492437					
Cranoglanis bouderius	Near	Near	Near			
Cromeria nilotica	Near	Near	Near			
Danio rerio	Near	Near	Near			
Denticeps clupeoides	Near		Near			
Diplomystes nahuelbutaensis	Near	Near	Near			
Distichodus notospilus	DQ492425					
Eigenmannia macrops	Near		Near			

Species	RAG1 Exon 3	ENC1	Plagl2	RAG2	CO1	*Cytb*
Electrophorus electricus	Near	Near	Near			
Erethistes sp. 1	DQ492449					
Erethistes sp. 2	DQ492450					
Erimyzon oblongus	Near	Near	Near			
Euchilichthys dybowskii	DQ492494			DQ492381		
Farlowella cf. nattereri	DQ492441					
Galeichthys ater	Novel	Novel	Near			
Galeichthys peruvianus	Near	Near	Near			
Glyptothorax cf. trilineatus	DQ492447					
Goeldiella eques	DQ492480					
Gogangra viridescens	DQ492448					
Gogo arcuatus	Near	Near	Near			
Gonorynchus abbreviatus	Near	Near	Near			
Gonorynchus greyi	Near	Near	Near			
Grasseichthys gabonensis	Near	Near	Near			
Gymnorhamphichthys petiti	Near		Near			
Gymnotus sp.	Near	Near	Near			
Helicophagus waandersii	DQ492515					
Helogenes marmoratus	DQ492534					
Hemibagrus wyckiodes	Near	Near	Near			
Hemisilurus moolenburghi	Near	Near	Near			
Henonemus punctatus	DQ492432					
Heterobagrus bocourti	Near	Near	Near			
Heterobranchus longifilis	DQ492520					
Heteropneustes fossilis	DQ492522					
Hoplias sp.	Near					
Hoplomyzon sexpapilostoma	DQ492536					
Horabagrus brachysoma	DQ492454					
Hydrolycus scomberoides	Near		Near			
Hypentelium nigricans	Near	Near	Near			
Hypophthalmus edentatus	DQ492474					

Species	RAG1 Exon 3	ENC1	Plagl2	RAG2	CO1	*Cytb*
Ictalurus punctatus	Near	Near	Near			
Ictiobus bubalus	Near	Near	Near			
Imparfinis cf. cochabambae	DQ492481					
Imparfinis cf. stictonotus	DQ492483					
Imparfinis stictonotus	DQ492482					
Ketengus sp.	DQ492526					
Kneria paucisquamata	Near	Near	Near			
Kneria ruaha	Near	Near	Near			
Kryptopterus minor	DQ492486					
Lacantunia enigmatica	EF078914			EF078916		
Laides hexanema	DQ492453					
Lamontichthys stibaros	DQ492440					
Leiocassis poecilopterus	Near	Near	Near			
Leporinus copelandii	Near	Near	Near			
Leptodoras linnelli	Near	Near	Near			
Liobagrus aequilabris	Near		Near			
Lophiobagrus aquilus	Novel	Novel	HG803292	HG803220	HG803449	HG803365
Lophiobagrus brevispinis	DQ492504		HG803291	HG803219	HG803448	HG803364
Lophiobagrus cyclurus	Novel		HG803295	HG803223	HG803452	HG803368
Lophiobagrus cyclurus			HG803307	HG803234	HG803464	HG803380
Lophiobagrus cyclurus	Novel		HG803312	HG803239	HG803471	HG803387
Loricaria simillima	Near		Near			
Malapterurus beninensis	Novel	Novel	Novel	Novel		
Malapterurus shirensis	Novel	Novel	Novel	Novel		
Malapterurus sp.	Novel	Novel				
Malapterurus tanganyikaensis	DQ492498			DQ492385		
Micromyzon akamai	DQ492537					
Microsynodontis sp.	DQ492496			DQ492383		
Mochokus niloticus	Novel	Novel	Novel	HF565739	HF565847	HF565995
Moxostoma macrolepidotum		Near	Near			
Mystus bimaculatus	Near	Near	Near			
Nematogenys inermis	Near	Near	Near			
Neolebias philippei			Near			
Neosilurus ater	DQ492529					

Species	RAG1 Exon 3	ENC1	Plagl2	RAG2	CO1	*Cytb*
Notemigonus crysoleucas	Near	Near	Near			
Noturus insignis	DQ492513					
Noturus stigmosus	Near	Near	Near			
Ochmacanthus alternus	DQ492433					
Olyra longicaudata	DQ492459					
Opsariichthyus uncirostris	Near	Near	Near			
Pangasianodon hypophthalmus	Near	Near	Near			
Pangasius larnaudii	DQ492516					
Parailia congica	Novel	Novel	HG803269	HG803196	HG803421	
Parailia sp.	DQ492509			DQ492396		
Parakneria slekii	Near	Near	Near			
Parakneria vilhenae	Near	Near	Near			
Parauchenoglanis balayi	DQ492500			DQ492387		
Parauchenoglanis fasciatus	Novel	Novel		HG803252	HG803488	HG803404
Parauchenoglanis ngamensis		Novel	HG803262	HG803190		HG803330
Pareutropius debauwi	Novel	Novel	HG803270	HG803197	HG803422	HG803338
Phalacronotus apogon	DQ492485					
Phenacogrammus interruptus	Near		Near			
Phractocephalus hemi	Near	Near	Near			
Phractolaemus ansorgii	Near	Near	Near			
Phractura lindica		Novel	Novel	Novel		
Phractura longicauda	DQ492490			DQ492377		
Phyllonemus aff. brichardi	Novel		HG803278	HG803205	HG803434	HG803350
Phyllonemus filinemus	Novel	Novel	HG803310	HG803237	HG803469	HG803385
Phyllonemus sp. A			HG803263	HG803191	HG803414	HG803331
Phyllonemus sp. B			HG803311	HG803238	HG803470	HG803386
Phyllonemus sp. C	Novel		HG803283	HG803211	HG803440	HG803356
Phyllonemus typus	DQ492503		HG803281	HG803209	HG803438	HG803354
Pimelodella cristata	DQ492478					
Pimelodus ornatus	DQ492475					
Plotosus lineatus	Near	Near	Near			
Porochilus rendahli	DQ492530					
Pseudeutropius brachypopterus	DQ492455					
Pseudopimelodus bufonius	DQ492471					

Species	RAG1 Exon 3	ENC1	Plagl2	RAG2	CO1	*Cytb*
Pseudopimelodus mangurus	DQ492472					
Pterobunocephalus sp.	DQ492535	Novel	Novel			
Pterocryptis anomala	DQ492487					
Pterygoplichthys multiradiatus	DQ492443					
Pygocentrus nattereri	Near	Near	Near			
Pylodictis olivaris	Near	Near	Near			
Rhamdia sp.	DQ492479					
Rhamphichthys sp.	Near	Near	Near			
Rheoglanis dendrophorus	Near	Near	Near	DQ492393		
Rita rita	DQ492518					
Schilbe intermedius	Novel	Novel	HG803277	HG803203	HG803432	HG803348
Scoloplax distolothrix	DQ492435					
Semotilus atromaculatus	Near	Near	Near			
Sternopygus macrurus	Near	Near	Near			
Sternopygus sp.	DQ492426					
Synodontis aff. ilebrevis	Novel	Novel	Novel		HF565878	DQ886644
Synodontis aff. schall	Novel	Novel	Novel	HF565817	HF565952	HF566067
Synodontis aff. tanganyicae	Novel	Novel	Novel	HF565831	HF565975	DQ886658
Synodontis afrofischeri	Novel	Novel	Novel	HF565744	HF565852	DQ886618
Synodontis angelica	Novel	Novel	Novel	Novel	HF565856	DQ886605
Synodontis batesii	Novel			HF565752	HF565862	HF566005
Synodontis grandiops		Novel	Novel		HF565890	FM878846
Synodontis granulosa	Novel	Novel	Novel	HF565777	HF565892	HF565777
Synodontis greshoffi	Novel	Novel	Novel		HF565894	HF566025
Synodontis irsacae	Novel	Novel	Novel	HF565767	HF565879	DQ886653
Synodontis lucipinnis	Novel	Novel	Novel	HF565787	HF565904	DQ886631
Synodontis membranaceus		Novel	Novel	HF565790	HF565908	HF566035
Synodontis multipuntata	Novel	Novel		HF565791	HF565910	DQ886625
Synodontis petricola	Novel	Novel	Novel			
Synodontis polli	Novel	Novel	Novel	HF565809	HF565941	DQ886645
Synodontis sorex	Novel	Novel	Novel	HF565823	HF565960	HF566074
Synodontis velifer	Novel	Novel	Novel	HF565836	HF565982	HF566089
Synodontis victoriae	Novel	Novel	Novel	HF565837	HF565984	DQ886657
Synodontis wamiensis	Novel	Novel	Novel	HF565839	HF565986	HF566092
Synodontis zambezensis	Novel	Novel	Novel	HF565844	HF565991	FM878858

Species	RAG1 Exon 3	ENC1	Plagl2	RAG2	CO1	*Cytb*
Trachelyopterus galeatus	DQ492464					
Tribolodon brandti	Near	Near	Near			
Trichomycterus guianense	DQ492431					
Wallago sp.	DQ492488					
Zaireichthys brevis	Novel	Novel	Novel	Novel		
Zaireichthys sp.	Novel	Novel		Novel		
Zaireichthys sp.	Novel			Novel		

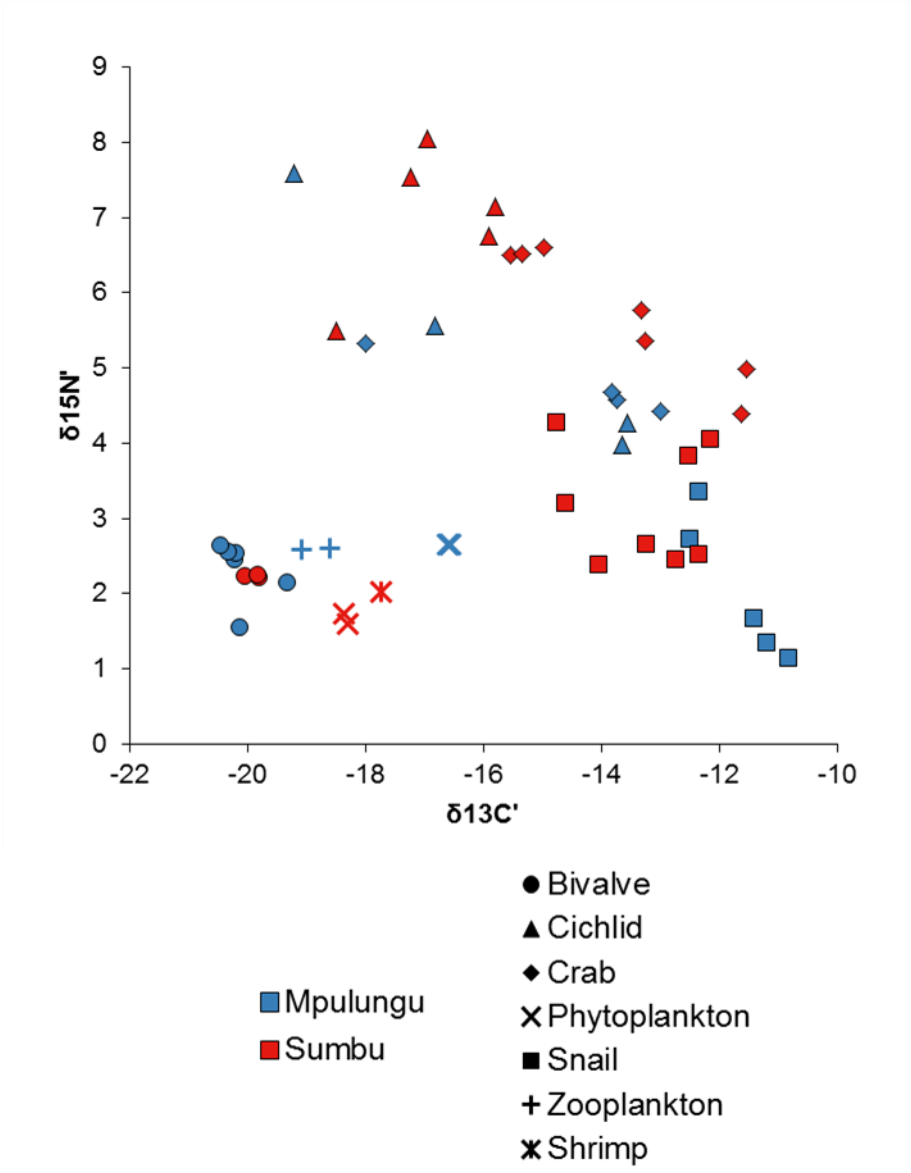
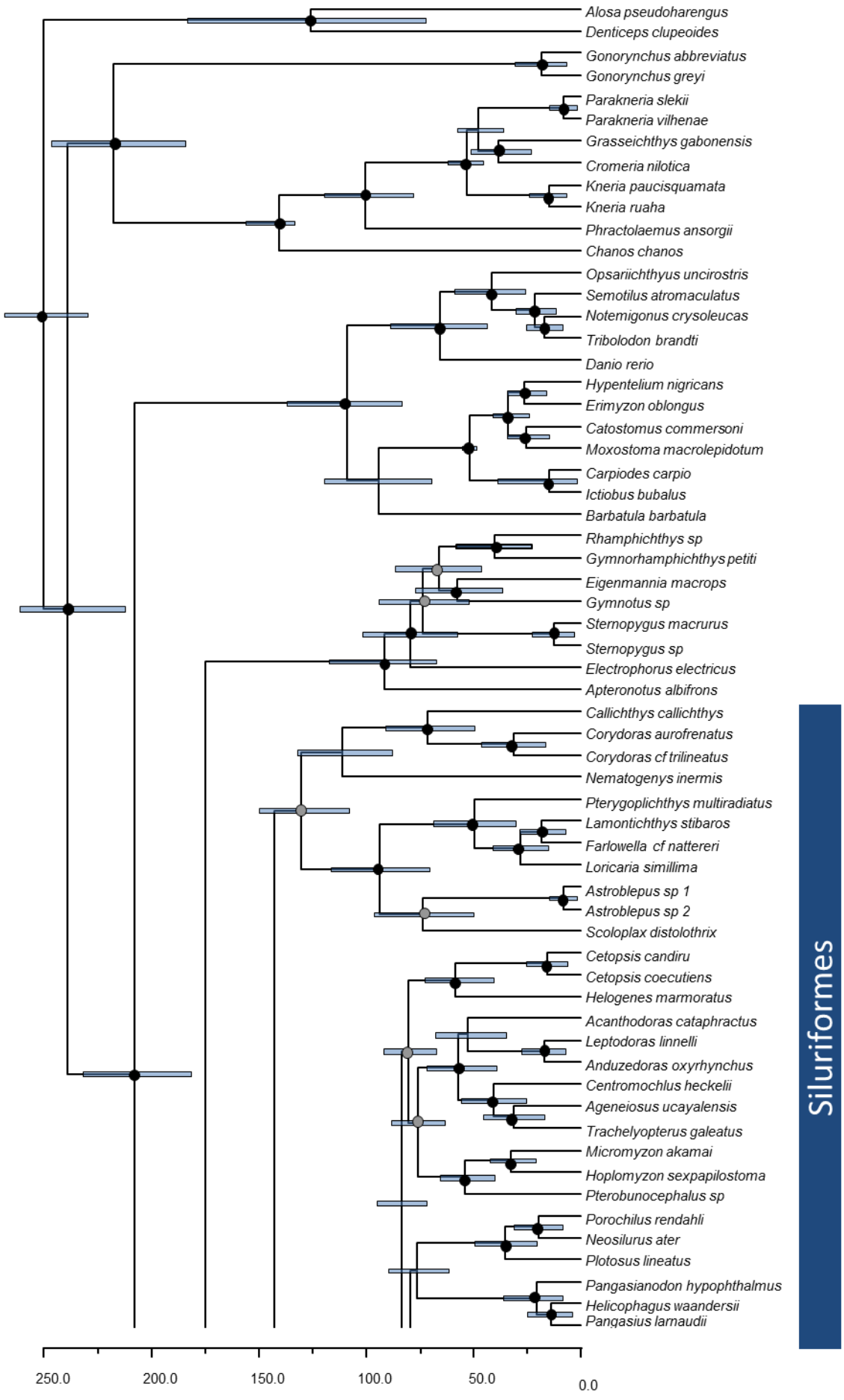
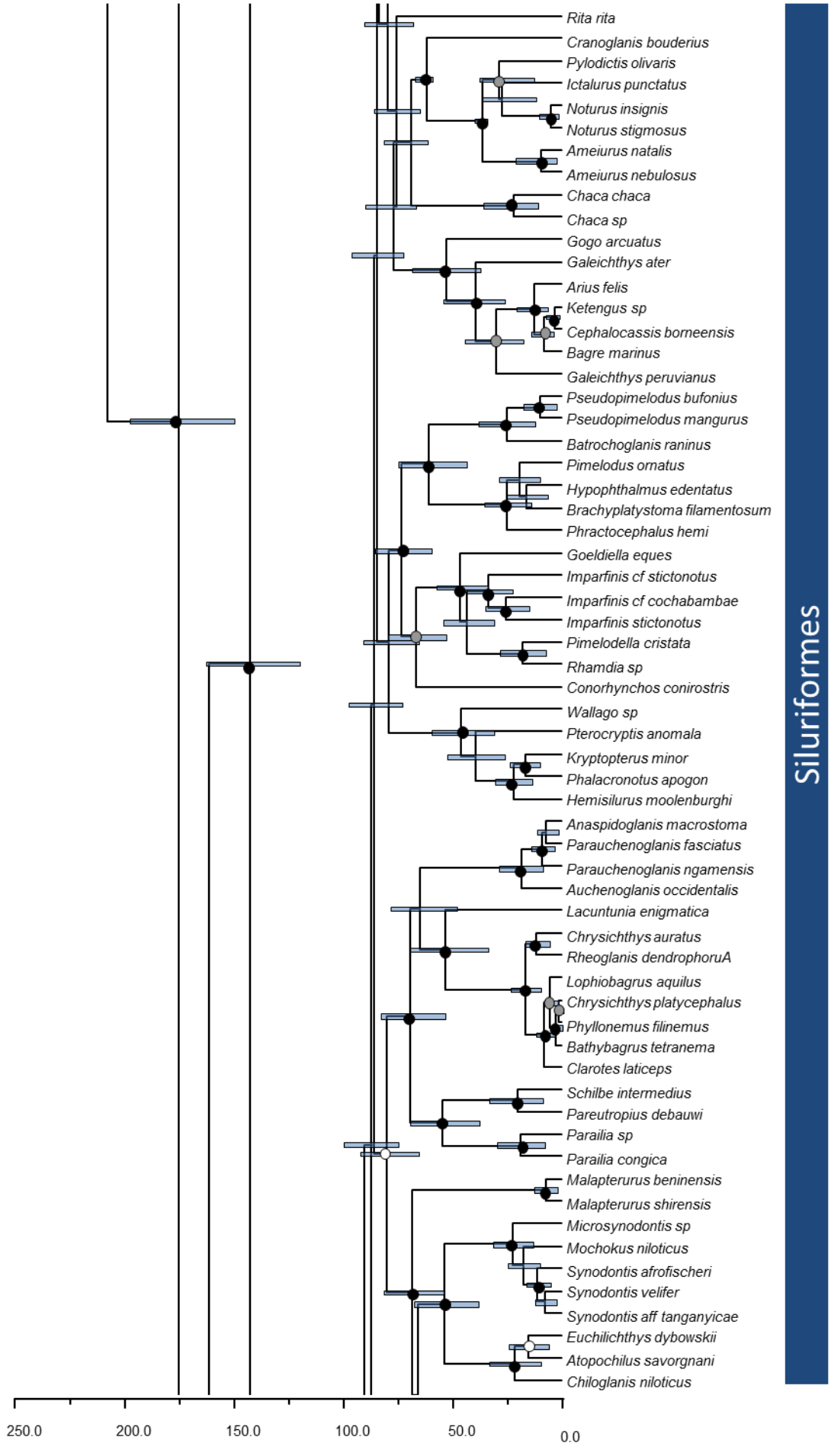


Figure 2 Scatter chart showing δ13C' vs δ15N' for baseline samples from Mpulungu (blue) and Sumbu (red). The baseline organisms are shown using different shaped markers outlined in the key.


Alosa pseudoharengus
Denticeps clupeoides
Gonorynchus abbreviatus
Gonorynchus greyi
Parakneria slekii
Parakneria vilhenae
Grasseichthys gabonensis
Cromeria nilotica
Kneria paucisquamata
Kneria ruaha
Phractolaemus ansorgii
Chanos chanos
Opsariichthyus uncirostris
Semotilus atromaculatus
Notemigonus crysoleucas
Tribolodon brandti
Danio rerio
Hypentelium nigricans
Erimyzon oblongus
Catostomus commersoni
Moxostoma macrolepidotum
Carpiodes carpio
Ictiobus bubalus
Barbatula barbatula
Rhamphichthys sp
Gymnorhamphichthys petiti
Eigenmannia macrops
Gymnotus sp
Sternopygus macrurus
Sternopygus sp
Electrophorus electricus
Apteronotus albifrons
Callichthys callichthys
Corydoras aurofrenatus
Corydoras cf trilineatus
Nematogenys inermis
Pterygoplichthys multiradiatus
Lamontichthys stibaros
Farlowella cf nattereri
Loricaria simillima
Astroblepus sp 1
Astroblepus sp 2
Scoloplax distolothrix
Cetopsis candiru
Cetopsis coecutiens
Helogenes marmoratus
Acanthodoras cataphractus
Leptodoras linnelli
Anduzedoras oxyrhynchus
Centromochlus heckelii
Ageneiosus ucayalensis
Trachelyopterus galeatus
Micromyzon akamai
Hoplomyzon sexpapilostoma
Pterobunocephalus sp
Porochilus rendahli
Neosilurus ater
Plotosus lineatus
Pangasianodon hypophthalmus
Helicophagus waandersii
Pangasius larnaudii
250.0
200.0
150.0
100.0
50.0
0.0
Siluriformes


Rita rita
Cranoglanis bouderius
Pylodictis olivaris
Ictalurus punctatus
Noturus insignis
Noturus stigmosus
Ameiurus natalis
Ameiurus nebulosus
Chaca chaca
Chaca sp
Gogo arcuatus
Galeichthys ater
Arius felis
Ketengus sp
Cephalocassis borneensis
Bagre marinus
Galeichthys peruvianus
Pseudopimelodus bufonius
Pseudopimelodus mangurus
Batrochoglanis raninus
Pimelodus ornatus
Hypophthalmus edentatus
Brachyplatystoma filamentosum
Phractocephalus hemi
Goeldiella eques
Imparfinis cf stictonotus
Imparfinis cf cochabambae
Imparfinis stictonotus
Pimelodella cristata
Rhamdia sp
Conorhynchos conirostris
Wallago sp
Pterocryptis anomala
Kryptopterus minor
Phalacronotus apogon
Hemisilurus moolenburghi
Anaspidoglanis macrostoma
Parauchenoglanis fasciatus
Parauchenoglanis ngamensis
Auchenoglanis occidentalis
Lacuntunia enigmatica
Chrysichthys auratus
RheoglanisdendrophoruA
Lophiobagrus aquilus
Chrysichthys platycephalus
Phyllonemus filinemus
Bathybagrus tetranema
Clarotes laticeps
Schilbe intermedius
Pareutropius debauwi
Parailia sp
Parailia congica
Malapterurus beninensis
Malapterurus shirensis
Microsynodontis sp
Mochokus niloticus
Synodontis afrofischeri
Synodontis velifer
Synodontis aff tanganyicae
Euchilichthys dybowskii
Atopochilus savorgnani
Chiloglanis niloticus
Siluriformes
250.0
200.0
150.0
100.0
50.0
0.0

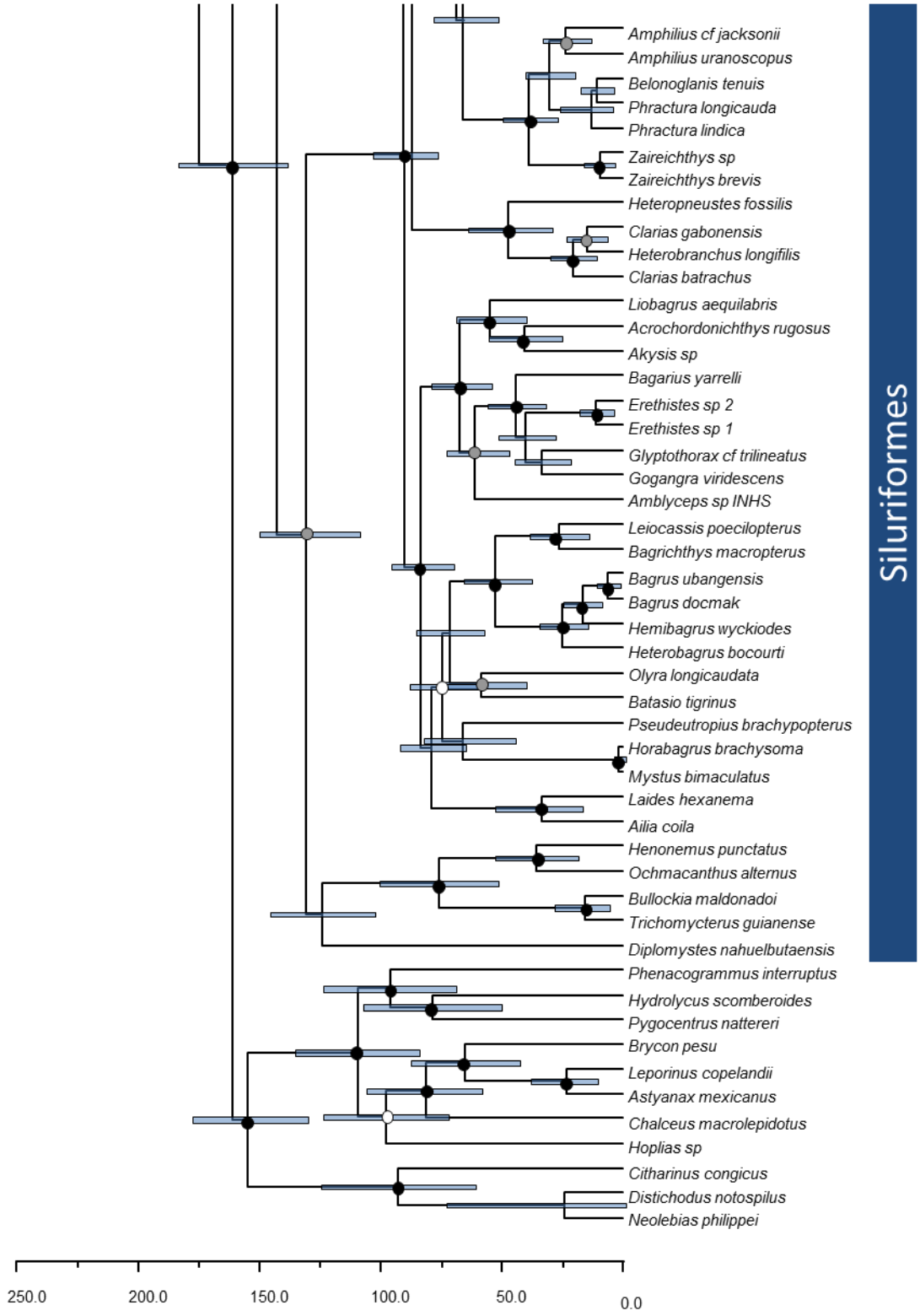


Figure 3 BEAST tree for the superorder Ostariophysi. The order Siluriformes (catfishes) is marked. Scale bar is Ma. Node bars show 95% confidence intervals around node ages. Black circles on nodes represent a posterior probability of 1, grey circles a posterior probability greater than 0.95, and white circles greater than 0.9.

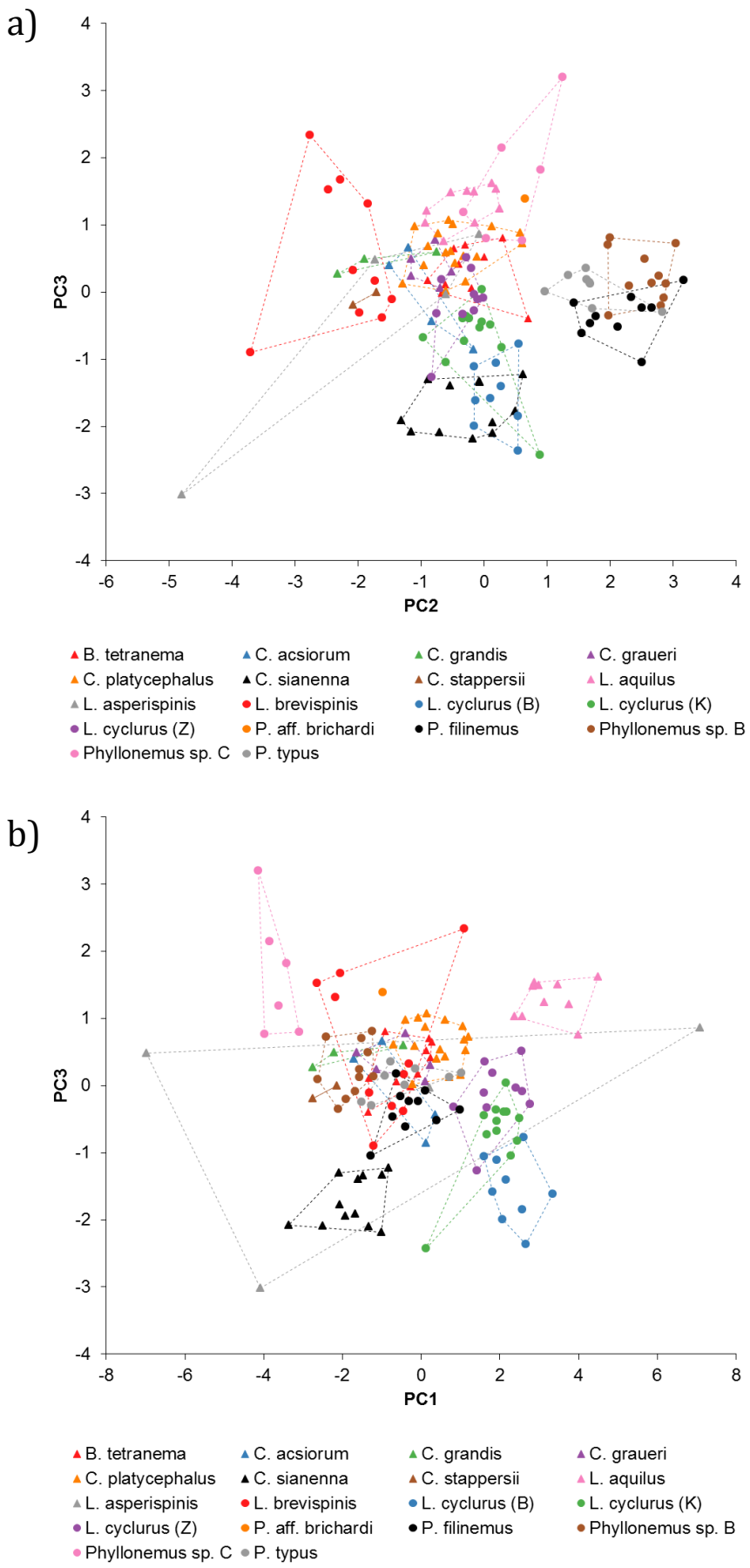
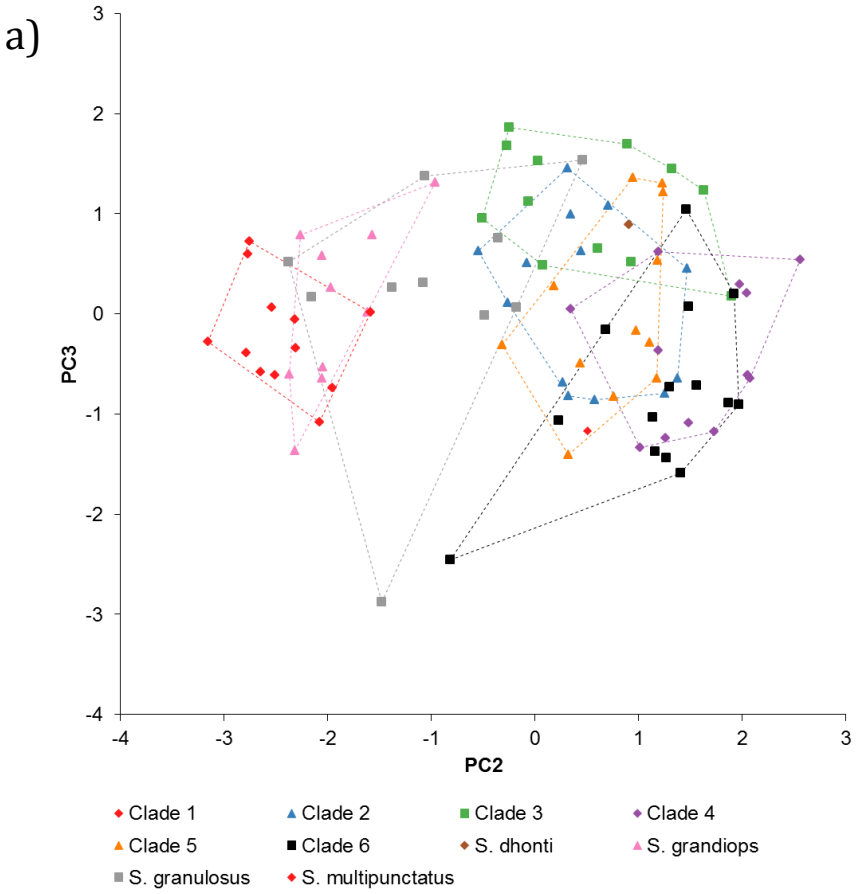


Figure 4 Non-phylogenetically corrected PCA plots for the claroteines, for a) PC2 vs PC3 and b) PC1 vs PC3.



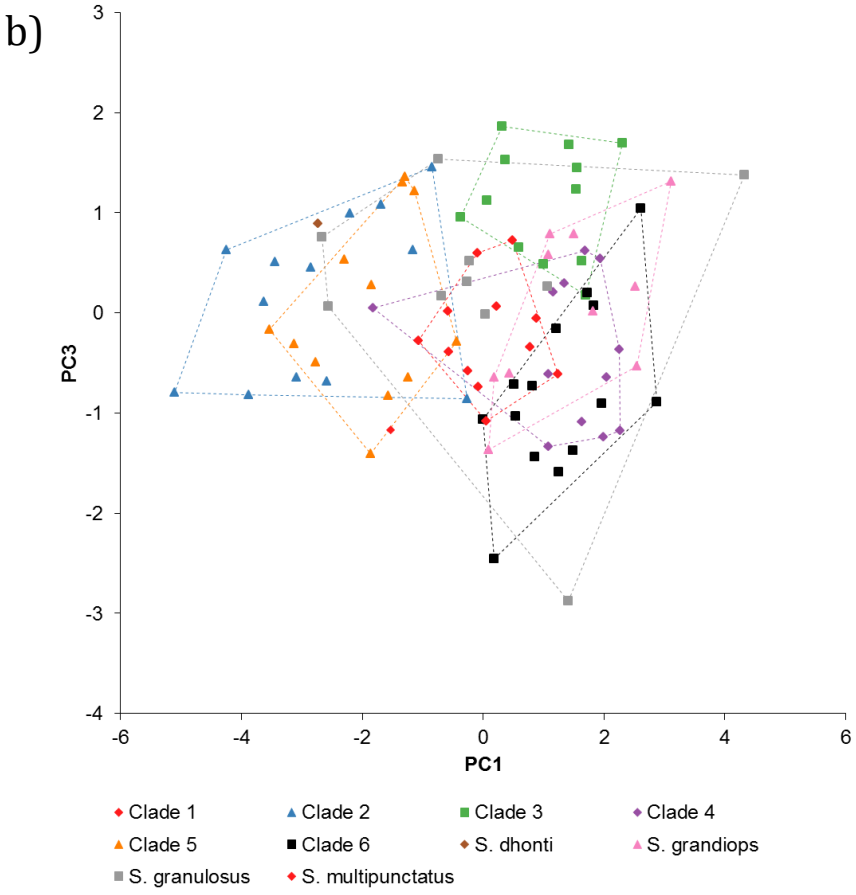


Figure 5 Non-phylogenetically corrected PCA plots for *Synodontis*, for a) PC2 vs PC3 and b) PC1 vs PC3.

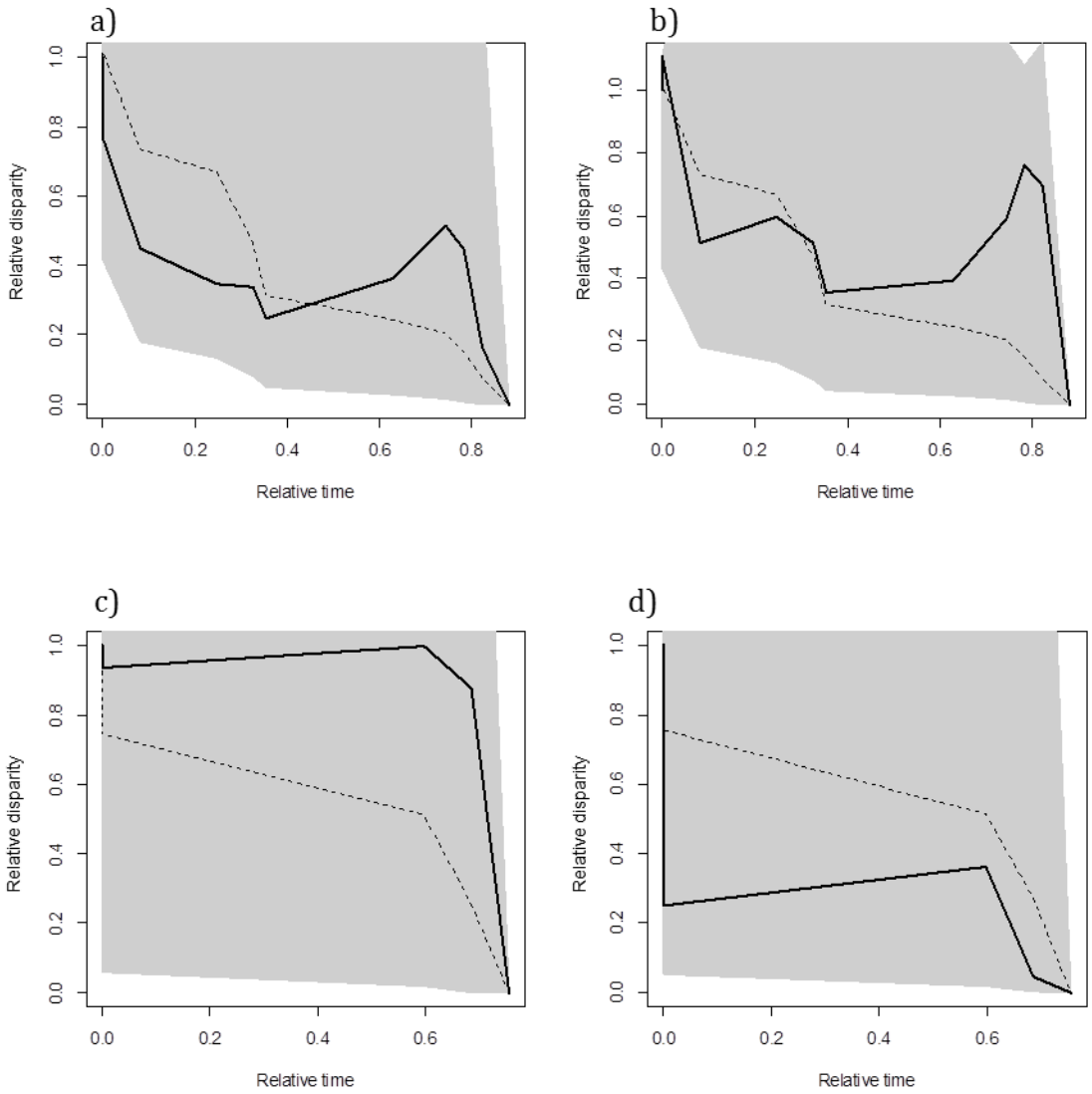


Figure 6 Stable isotope disparity through time plots for the claroteines (a – Carbon, b – Nitrogen) and Synodontis (c – Carbon, d – Nitrogen).

Table 3 MDI for δC^{13} and δN^{15} for the claroteines using the MCC tree and 1000 trees.

	MCC tree		1000 Trees			
	MDI	P-value	MDI mean	MDI standard deviation	P-value mean	P-value standard deviation
δC^{13}	-0.038	0.357	-0.041	0.040	0.358	0.086
δN^{15}	0.088	0.373	0.070	0.043	0.399	0.071

Appendix 4

Peart, C. R., Bills, R., Wilkinson, M., & Day, J. J. (2014) Nocturnal claroteine catfishes reveal dual colonisation but a single radiation in Lake Tanganyika. *Molecular Phylogenetics and Evolution*, **73**, 119–128